

Functional modelling of interaural time difference discrimination in acoustical and electrical hearing

Andreas Prokopiou, Arturo Moncada-Torres, Jan Wouters, Tom Francart
June 12, 2017

Abstract

Objective

Interaural time differences (ITDs) are important for sound source localisation. We present a model to predict the just noticeable differences (JNDs) in ITD discrimination for normal hearing and electric stimulation through a cochlear implant.

Approach

We combined periphery models of acoustic and electric stimulation with a novel JND in ITD estimation stage, which consists of a shuffled cross correlogram and a binary classifier characterisation method. Furthermore, an evaluation framework is presented based on a large behavioural dataset.

Main results

The model correctly predicts behavioural observations for unmodulated stimuli (such as pure tones and electric pulse trains) and modulated stimuli for modulation frequencies below 30 Hz. For higher modulation frequencies, the model predicts the observed behavioural trends, but tends to estimate higher ITD sensitivity.

Significance

The presented model can be used to investigate the implications of modifying the stimulus waveform on ITD sensitivity, and as such be applied in investigating sound encoding strategies.

1 Introduction

Cochlear implants (CI) are prosthetic devices that can restore hearing to profoundly deaf persons. Essentially, they consist of a receiver stage and a signal processing stage before electrically stimulating the spiral ganglion using multiple electrodes. Implanting patients with CIs on both ears (i.e., bilateral implantation) is becoming a common clinical practice, particularly in young children. It has been shown that it improves sound localization and speech perception in noise over unilateral CIs (e.g., Firszt et al., 2008; Laback et al., 2015; Offeciers et al., 2005; Van Deun et al., 2009). These benefits arise mostly from utilizing the acoustic head shadow effect, which creates an interaural level difference (ILD) between the two ears (Van Hoesel, 2012). Another important cue for sound localization in normal hearing (NH) listeners is the interaural time difference (ITD), which is the most salient cue for sounds with sufficient low-frequency content (Macpherson and Middlebrooks, 2002; Wightman and Kistler, 1992). Furthermore, ITDs have been shown to be dominant for

both binaural unmasking and attention-driven spatial release from masking (Bronkhorst, 2000; Bronkhorst and Plomp, 1988; Kidd et al., 2010).

Unfortunately, ITD cues are not fully coded by clinical CI sound processors. Furthermore, bilateral cochlear implant (BiCI) users have been shown to have poor sensitivity to high-rate pulse train ITDs (Laback et al., 2015; Noel and Eddington, 2013). However, behavioural studies have given evidence to support the idea that modifying the modulating envelope of a carrier tone has the potential to influence ITD perception in BiCI users (Laback et al., 2011; Noel and Eddington, 2013; Van Hoesel et al., 2009) and in bimodal users (i.e. one ear with a hearing aid and the other with a CI, Francart et al., 2009, 2011, 2014). There are still unknowns about ITD perception when using a CI. Specifically, how do temporal properties of the stimulation envelope (such as modulation frequency and depth), temporal gaps, and the rate of change of envelope amplitude affect ITD perception.

Various binaural models have been developed which attempt to address these questions by describing the complex interactions between perceived location due to ITD and stimulus waveform. These models can be classified into two categories. On one hand, *biophysical models* aim to describe particular neural pathways by characterising underlying binaural mechanisms (Chung et al., 2015; Rothman and Manis, 2003; Wang and Colburn, 2012; Wang et al., 2014). On the other hand, *statistical models* (also known as functional or phenomenological models) aim to describe a generic binaural processor using signal processing techniques that relate empirical observations to the model outcomes (Bernstein and Trahiotis, 2002, 2009; Breebaart et al., 2001; Colburn, 1973; Colburn and Durlach, 1978; Dietz et al., 2011; Pulkki and Hirvonen, 2009; Takanen et al., 2014). However, this type of models typically do not attempt to explain the underlying mechanisms behind said observations.

Furthermore, a substantial body of work already exists for describing the acoustic and the electric stimulation of the auditory nerve (AN). In the acoustic case, the model proposed by Zilany et al. (2009) for acoustical stimulation can accurately predict various temporal phenomena such as non-linear tuning, level-dependent phase, compression, suppression, shift in the best frequency as a function of level, adaptation, as well as some other non-linearities seen at high sound levels based on several AN datasets. Additionally, its most recent version corrected the saturation of firing rates of higher characteristic frequency fibres when stimulated by low frequency tones (Zilany et al., 2014). This modelling effort improved the prediction of AN responses to a wide variety of complex sounds (such as amplitude-modulated stimuli) and forward-masking paradigms (Zilany et al., 2009), while accounting for long-term dynamics of AN responses (Zilany and Carney, 2010).

In the electric case, the situation is rather different. There is a broader landscape of models which can be separated into three main categories (Nicoletti et al., 2013). *Point neuron models* (Bruce et al., 1999; Goldwyn et al., 2012; Mino et al., 2002; Motz and Rattay, 1986; Rattay, 1986) aim at modelling individual neuron detailed dynamic properties. *Multi-compartmental models* (Briaire and Frijns, 2000; McNeal, 1976; Mino et al., 2004; Woo et al., 2010) work as extensions of point neuron models by considering how and where the action potentials are generated on the AN following electrical stimulation. They are typically used for connecting a sequence of neurons. Finally, *Population models* (Nicoletti et al., 2010) aim at replicating excitation patterns along the entire cochlea.

The complex relation between stimulus waveform and ITD perception is not yet fully understood. In order to describe human ITD perception, we developed a computational model relying on the working hypothesis that central auditory processing is normal (i.e.,

not impaired) for BiCI users, which is also adopted by Chung et al. (2015). As such, a single model is used to process the AN response for both NH and BiCI users by utilizing their respective acoustic and electric stimuli at AN level as inputs. For the acoustic case we chose the aforementioned model proposed by Zilany et al. (2014) to represent the neural responses of the acoustic stimulation peripheral pathway. For the electric case, we considered direct stimulation of the spiral ganglion without any residual hearing. Furthermore, we considered a single bilateral electrode pair stimulation in order to reduce the number of assumptions that would be necessary to account for across channel integration phenomena. Thus, a point neuron approach is sufficient, since capturing temporal details of the dynamic response of the AN is vital to single electrode pair stimulation. In particular, we chose the model proposed by Goldwyn et al. (2012), which utilises point-process analysis to modify a parameter space to tune neuron firing parameters, such as response latency, threshold, relative spread, jitter, and summation time. The model includes temporal filtering that represents sub-threshold dynamics of the membrane potential, a non-linearity associated with spike generating processes, and a secondary filter that accounts for variability in spike timing. It incorporates dynamical and stochastic properties that are important to high pulse rate stimulation, which is particularly relevant to the clinical stimuli used in modern CI devices that can reach stimulation rates in the range of 1,000 - 20,000 pulses per second (PPS). Both the acoustic and electric front end were combined with a novel neurometric psychometric estimation method. We validated the model by comparing its just noticeable difference (JND) in ITD predictions with relevant psychoacoustic experimental results from the literature.

2 Methods

The framework of the proposed computational model is illustrated in Fig. 1. A particular pair of left and right ear stimuli were given as input to a model of the human auditory periphery. The resulting output of the periphery described the neuronal activity of the AN for either acoustic or electric stimulation. This neuronal activity was quantified as a temporal pattern of action potentials and was used as an input to the next stage of the model: the shuffled cross correlogram (SCC, Joris et al., 2006). This stage gave an estimate of the relative timing disparity between the left and the right AN activity.

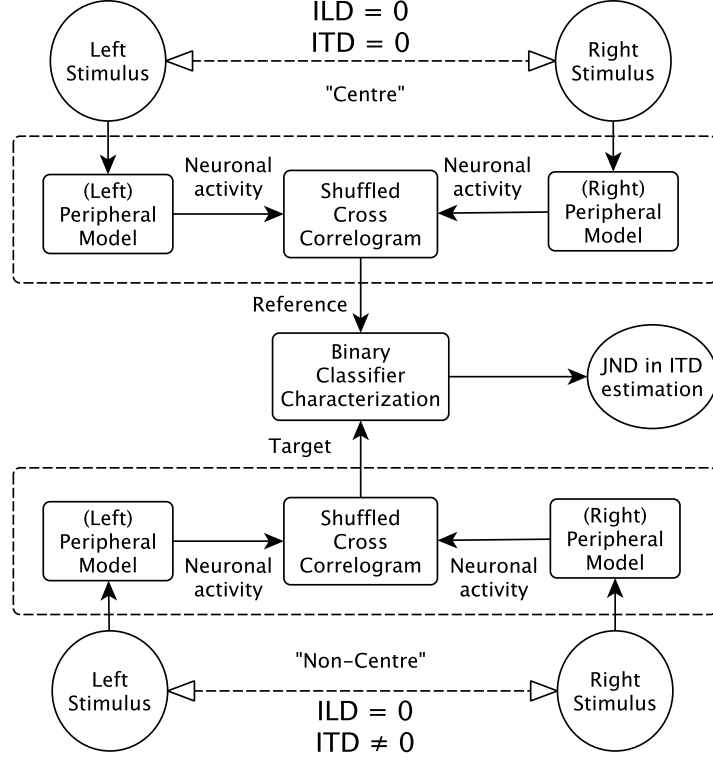


Figure 1: Overview of the model framework. Note that the blocks in the dashed box are identical. $ITD \neq 0$ refers to an example case, since the ITD is a model parameter that is varied to estimate the JND in ITD.

The left and right ear stimuli used as inputs to the model were either presented with an ITD equal to 0, serving as a reference, or with an ITD that is non-zero, serving as a target. Both the target and the reference were used as input to a binary classifier characterization stage where the JND in ITD was estimated.

The model parameters are the neural density distribution on the basilar membrane, the bin-width of the SCC, and jitter introduced after the peripheral neuronal activity estimation. The remaining parameters particular to the peripheral models were adopted from their respective publications (Goldwyn et al., 2012; Zilany et al., 2014). Specifically, for the acoustic stimulation the parameters of the Zilany et al. (2014) model are the functioning of the outer and inner hair cells, which was set to normal, i.e. healthy function; the species model, which was set to be human with basilar membrane tuning from Shera et al. (2002); the AN spontaneous rate which was set to be a high spontaneous rate; the noise type which was set to be variable, i.e. different every simulation, and the implementation of the power-law function which set to be the actual implementation (i.e. not an approximation). Furthermore, the binsize for the resulting PSTH was set to be equal to the sampling time to get the individual timing for each action potential. The individual time is important for the calculation of the SCC (Sec. 2.3). For the electric stimulation the parameters of the Goldwyn et al. (2012) model are a threshold of 0.852 mA (Miller et al., 2001), a relative spread of 4.87% (Miller et al., 2001) a jitter of $85.5 \mu\text{s}$ (Miller et al., 2001) and a summation time of $250 \mu\text{s}$ (Cartee et al., 2006).

2.1 Stimuli

The proposed computational model aims to estimate ITD sensitivity for both acoustic and CI stimulation. All the stimuli were previously described in various behavioural studies where the JND in ITD was reported. The selection of these studies aims to create an evaluation framework for the model.

2.1.1 Acoustic stimuli

For the NH situation, the subject was assumed to have normal acoustic hearing on both ears. The setup of the model permitted estimation of JND in ITD for situations where the audio is narrowband. The peripheral model operated within a bandwidth of one equivalent rectangular bandwidth (ERB, Glasberg and Moore, 1990). Corresponding to the behavioural data that was used, the ITD was static and there was no ILD. Furthermore, the behavioural studies considered ongoing cues, so the onset and offset cues were reduced using a 20 ms cosine square ramp for all acoustic stimuli.

The model's predictions were compared against behavioural data for pure tones (Brughera et al., 2013), sinusoidally amplitude modulated (SAM) tones (Bernstein and Trahiotis, 2002), transposed tones (Bernstein and Trahiotis, 2002), raised sine tones both as the modulation frequency and the modulation depth is varied (Bernstein and Trahiotis, 2009), and finally as the tone is modulated with trapezoidal shaped waveforms (Laback et al., 2011). The mathematical definition of the stimuli is given in the Appendix. All the acoustic stimuli parameters are shown in Table 1.

	f_c (kHz)	f_m (Hz)	m	D (ms)	Intensity (dB SPL)	Extra variable(s)
Pure tones	0.25, 0.5, 0.7, 0.8, 0.9, 1, 1.2, 1.25, 1.3	N/A	N/A	300	75	N/A
SAM tones	4, 6, 10	30, 60, 120, 250, 500	1	300	75	N/A
Transposed	4, 6, 10	30, 60, 120, 250, 500	1	300	75	N/A
Raised sine (f_m)	4	30, 60, 120, 250, 500	1	300	75	exponent (n) =... 1, 2, 4, 8
Raised sine (m)	4	128	0.25, 0.5, 0.75, 1	300	75	exponent (n) = ... 1, 2, 4, 8 slope = ...
Trapezoid	8.727	27.3	1	1000	78.2	6, 8, 10, 12 dBms ⁻¹ off time = 1, 6, 12, 18, 21 ms

Table 1: NH stimulus parameters. All stimuli were generated with a sampling frequency of 100 kHz. Abbreviations are: carrier frequency (f_c), modulation frequency (f_m), modulation depth (m), duration (D), sinusoidally amplitude modulated (SAM).

2.1.2 Electrical stimuli

For the BiCI situation, the subject was assumed to have no residual hearing and a CI implanted in both ears. The electric pulses for all the electric stimuli used were biphasic

with a phase duration of 25 μ s and an interphase gap of 8 μ s. The pulse rate was variable, ranging from 40 to 5000 PPS (depending on the behavioural data). Defining the stimulation current amplitude is not a straightforward task as the connection between loudness perception and stimulating current is highly subject dependent. In order to quantify the CI stimulation intensity, we used the firing efficiency measure, which is explained as follows.

The firing efficiency curve is an input/output function that relates the current level of a single pulse of current to the probability that the stimulus evokes a spike. Typically it resembles a sigmoid curve (see Appendix). The definition of threshold current (I_{thr}) for a particular neuron is when the probability of generating an action potential is equal to 50%. The simulated neuron had $I_{thr} = 0.852$ mA. The peak current, I_{peak} , for each experimental condition was set to have a firing efficiency of 1.25 dB with reference to I_{thr} , which means that $I_{peak} = I_{thr} * 10^{\frac{1.25-1}{20}} = 0.877$ mA. Using the firing efficiency curve this I_{peak} value corresponded to approximately a 75% chance of generating an action potential from a single pulse. The stimulation current in all the electric stimuli used was scaled to have the maximum current be equal to the calculated peak current.

The firing efficiency of 1.25 dB was chosen because the average firing rate of neurons activated with electric stimulation was found to be comparable to acoustic stimulation at 75 dB SPL. The comparison method between electric and acoustic stimulation consisted of estimating the baseline firing rate of either, and varying the firing efficiency until the baseline firing rate matched. The baseline firing rate was estimated as a running average of spikes per neuron over a 15 ms time window for low rate stimulation of constant envelope (i.e. unmodulated) stimuli.

The binaural model was compared against behavioural data for ITD sensitivity of unmodulated low-frequency pulse trains (Egger et al., 2016; Laback et al., 2007; van Hoesel and Clark, 1997; van Hoesel and Hoesel, 2007; Van Hoesel et al., 2009; van Hoesel and Tyler, 2003), SAM modulated pulse trains (Noel and Eddington, 2013) and trapezoidally modulated pulse trains (Laback et al., 2011). All the electrical stimuli parameters are shown in the Table 2.

	Pulse rate (PPS)	f_m (Hz)	m	D (ms)	Firing efficiency (dB)	Extra variable(s)
Pulse train	40, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000	N/A	N/A	300	1.25	N/A
SAM pulse train	1000 5000	4, 8, 16, 50, 100, 200, 500	1	300	1.25	N/A
Trapezoid	1515	27.3	1	1000	1.25	slope = ... 6, 8, 10, 12 dBms ⁻¹ off time = 1, 6, 12, 18, 21 ms

Table 2: Bilateral CI stimulus parameters. All stimuli were generated with a sampling frequency of 100 kHz. Abbreviations are: pulses per second (PPS), modulation frequency (f_m), modulation depth (m), duration (D) and sinusoidally amplitude modulated (SAM).

2.2 Peripheral model

The peripheral model stage emulated the physiological conversion of an external stimulus (either acoustic or electric), to a binary sequence of action potentials on the AN. We used two identical peripheral models for the left and the right ears.

The number of acoustical neurons simulated (N_{acou}) was set equal to the amount of neurons that spanned one equivalent rectangular bandwidth (ERB), as described by Glasberg and Moore (1990) and defined in Eq. (1). The centre frequency (CF) of the band was chosen to be equal to the carrier frequency of the stimulating tone.

$$ERB = 24.7 \left(\frac{4.37CF}{1000} + 1 \right)$$

$$f_{low} = CF - \frac{ERB}{2}, f_{high} = CF + \frac{ERB}{2} \quad (1)$$

The calculation of N_{acou} is shown in Eq. (2), where f_{high} and f_{low} are the frequency bounds of the ERB and x_{low} and x_{high} are the corresponding locations on the basilar membrane of humans as described by Greenwood (1990) and defined in Eq. (2). Furthermore, the density of neural innervation of the human basilar membrane was defined as $\rho(x)$, where x is distance along the basilar membrane, as described by Spoendlin and Schrott (1988). Specifically, $\rho(x)$ was calculated using a polynomial fit to their data to interpolate the measurements required by the model. The equivalent consideration for the electric case is the spread of excitation. However, it was not considered in this work because only one electrode was simulated in each cochlea. The number of electrically stimulated neurons was fixed at $N_{elec} = 1000$ neurons, which was similar on average to N_{acou} (i.e., the amount of neurons that roughly span one ERB).

$$x_{low} = \frac{1}{2.1} \log_{10} \left(\frac{f_{low}}{165.4} + 0.88 \right), x_{high} = \frac{1}{2.1} \log_{10} \left(\frac{f_{high}}{165.4} + 0.88 \right)$$

$$N_{acou} = \int_{x_{low}}^{x_{high}} \rho(x) dx \quad (2)$$

The output of the action potential estimating models (both acoustic and electric) was modified by introducing temporal jitter with a standard deviation of $250 \mu s$. This was done to simulate noise addition from action potential propagation (Faisal et al., 2008) and to remove the phase locking to the pulsatile electric stimulus to emphasise the envelope.

2.3 Shuffled Cross Correlogram

The Shuffled Cross Correlogram (SCC, Joris et al., 2006) is a variation of the Shuffled Auto Correlogram (SAC) proposed by Joris (2003). The SCC functions as a binaural coincidence detector by comparing the firing timing between neurons of the left and right auditory fibers (Fig. 2). It was computed as follows. First, N spike trains from the left ear and N spike trains from the right ear were fed as inputs. Then, the forward and backward time intervals between all the spikes of the first left spike train and all the spikes of all the right spike trains were measured. These time intervals were tallied into a histogram. For the latter, a standard binwidth size $\Delta\tau$ of $50 \mu s$ was used, as suggested

by Louage et al. (2004). A schematic representation of the SCC computation is shown in Fig. 2. The operation of counting intervals can be thought as an equivalent of counting coincident spikes between two different spike trains. Therefore, the SCC can be plotted as the number of intervals (or counts) versus the delay value (Louage et al., 2004).

Since we were interested in the firing temporal properties, the SCC was normalized by N , average firing rate r , $\Delta\tau$, and stimulus duration D . This was done by dividing the SCC by the term $N^2 r^2 \Delta\tau D$, making it dimensionless and independent from these parameters. Thus, a count value larger than 1 means that spikes tend to be correlated between spike trains; a count value of 1 shows a lack of stimulus-induced temporal correlation; a count value smaller than 1 indicates anticorrelation (Joris et al., 2006).

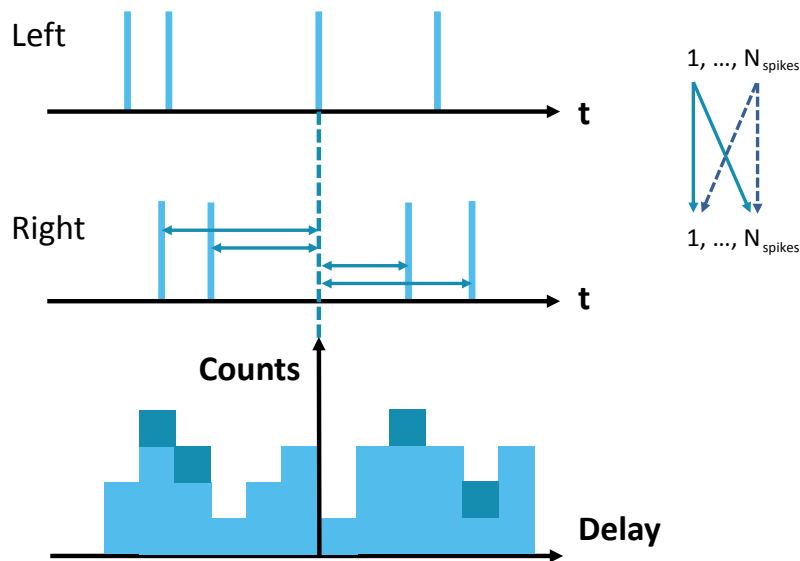


Figure 2: Schematic representation of the Shuffled Cross Correlogram (SCC) computation.

2.4 Binary classifier characterisation

The SCC yielded a distribution which characterises the joint temporal properties of the action potentials generated at both left and right ANs. The SCC distribution translates along the x-axis as the ITD between left and right stimuli varies (Fig. 3A). Therefore, the purpose of the binary classifier characterisation (BCC) method is to produce a metric which quantifies the mismatch between a reference distribution (ITD = 0, i.e., non-shifted) and target distribution (ITD \neq 0, i.e., shifted) in such a way so as to provide an estimation of JND in ITD.

The mismatch between the reference and the target distributions was quantified in an analogous way to the computation of the area under the curve (AUC) in a receiver operating characteristic (ROC) curve. There are some differences, which are explained as follows. Typically, when computing the ROC curve, the ordinate is the true positive rate and the abscissa is the false positive rate. In the BCC method the cumulative surface area of the reference distribution ($CS_R(\sigma)$), is the ordinate, and the cumulative surface area of the target distribution ($CS_T(\sigma, ITD)$) is the abscissa; where σ is the SCC lags. The cumulative surface areas of the two distributions are computed within an integration window w as shown in Fig. 3A.

The AUC is computed for various integration window widths and is defined in Eq. (3) as $\Delta A(w, ITD)$; where w corresponds to the width of the integration window of coinci-

dence detection neurons in the medial superior olive (MSO). An exploration of $\Delta A(w)$ as w changes shows that it resembles a dampened oscillation (Fig. 3B), where its maximum value ΔA_{max} , defines $\psi(ITD)$ as shown in Eq. (4).

$$\Delta A(w, ITD) = \int_{-w}^w CS_R(\sigma) \frac{\partial CS_T(\sigma, ITD)}{\partial \sigma} d\sigma \quad (3)$$

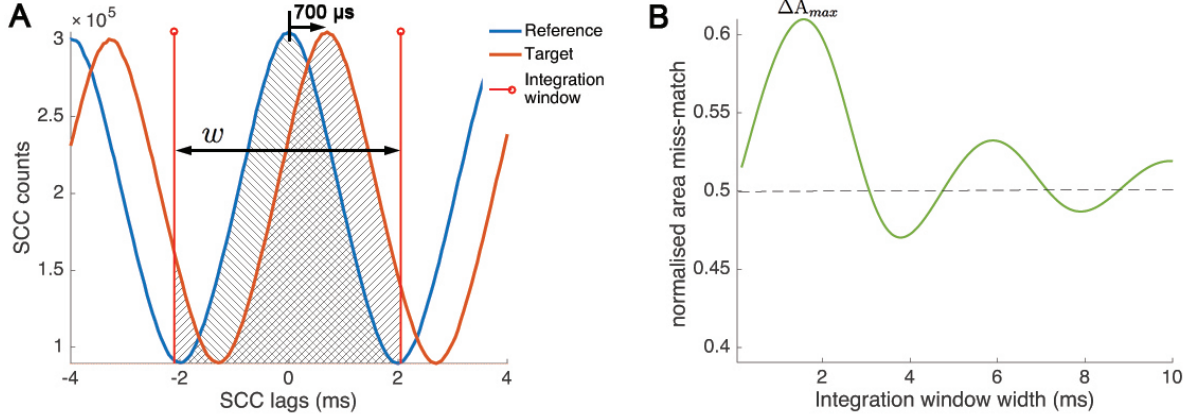


Figure 3: (A) SCC curve for a pure tone with frequency of 250 Hz at 75 dB SPL with a reference $ITD = 0$ and target $ITD = 700 \mu s$. The BCC method relates the difference of the shaded areas. (B) Plot of $\Delta A(w)$ when $ITD = 700 \mu s$, (Eq. (3)). It is a measure of the normalised area mismatch between the reference and target SCC of panel A. Note that the dashed line at 0.5 indicates the point where the two areas are equal.

$$\begin{aligned} w_{max} &= \operatorname{argmax}_w (|\Delta A(w, ITD)|) \\ \psi(ITD) &= \Delta A(w_{max}, ITD) \end{aligned} \quad (4)$$

The action of selecting the maximum value of $\Delta A(w, ITD)$ as w varies has the effect of eliminating the w variable by fixing it as the constant w_{max} shown in Eq. (4); hence $\psi(ITD)$ varies only as ITD changes for a particular stimulus. For example, Fig. 3B shows that $\psi \approx 0.61$ (labelled with ΔA_{max}) for a pure tone with frequency of 250 Hz and an $ITD = 700 \mu s$. Fig. 4A shows cases with different ITDs. As the ITD increases, so does the value of $\psi(ITD)$.

The $\psi(ITD)$ metric can be thought of as a neurometric-psychometric (NP) method which applies signal detection theory to map neural activity to stimulus sensation (Stüttgen et al., 2011). The dependence of $\psi(ITD)$ on ITD is illustrated in Fig. 4B, where we observe that $\psi(ITD)$ is a neurometric function derived from the BCC analysis. The $\psi(ITD)$ has certain properties which make it comparable to a psychometric function for the region of physiologically relevant ITD values. Specifically, when the reference and target distributions overlap $\psi(ITD) = 0.5$; it is normalised such that $0 < \psi(ITD) < 1$; and there is a logistic function whose inflection point lies on the threshold, i.e. when the reference and target distributions overlap at $ITD = 0$.

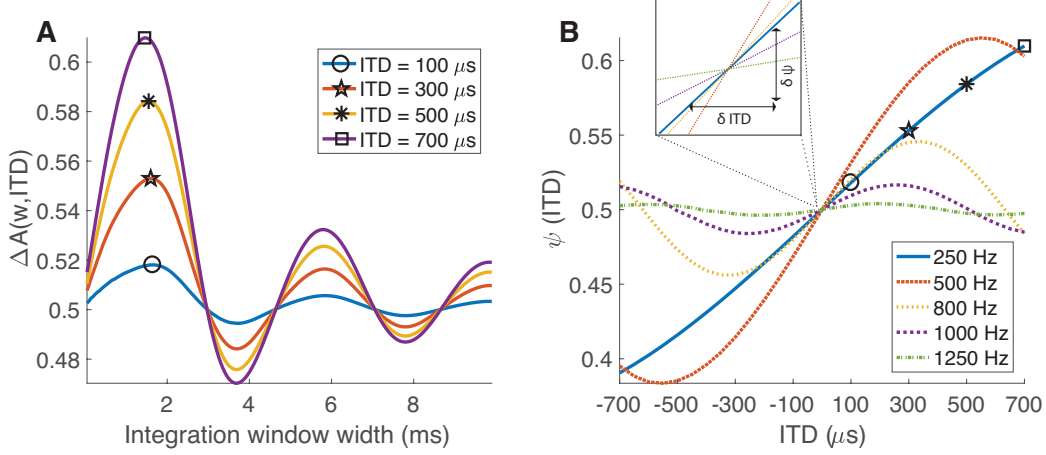


Figure 4: **(A)** Various $\Delta A(w, ITD)$ curves exemplifying the AUC computation (Eq. (3)) as a function of the integration window width (w) and various ITD values. The stimulus was a pure tone with frequency of 250 Hz at 75 dB SPL. The symbols ($\circ, \star, *, \square$) show the ΔA_{max} of the particular curve and correspond to the $\psi(ITD)$ value on panel **B**. **(B)** Neurometric-psychometric $\psi(ITD)$ curves of pure tone stimuli of various frequencies as a function of ITD. The insert shows the slope estimation at $ITD = 0$, where the slope is approximated as $\delta\psi/\delta ITD$.

The next step was to relate $\psi(ITD)$ with an estimate of the JND as shown in Eq. (5). Since $\psi(ITD)$ resembles a psychometric function, the JND was defined the inverse of the slope at $ITD = 0$ (insert of Fig. 4B). A common JND measure is defined by the non-zero value of a varied parameter (here the ITD) where the psychometric function reaches a certain (defined) performance level. The steeper the slope at the threshold (i.e. the inflection point of the logistic function) the less change in the varied parameter is needed (i.e. smaller JND) before the defined performance level is attained. As such, the slope is inversely proportional to the value of JND. This relation is described in Eq. (5). It produces a dimensionless measure of JND and thus receives the arbitrary quantity of *model units*.

$$JND(ITD) = \left(\frac{d\psi(ITD)}{dITD} \right)^{-1} \quad (5)$$

Note that for the JND estimations presented in Sec. 3, the reference $ITD = 0$ (Bernstein and Trahiotis, 2002, 2009; Brughera et al., 2013; Laback et al., 2011). Thus we define the JND as the minimum detectable change, or ITD threshold, for a reference $ITD = 0$.

2.5 Model performance evaluation

The linear correlation (Pearson's r) between model and behavioural data was calculated to objectively compare model predictions and behavioural data for each dataset. Linear scale was used to order the model and the behavioural data, in respectively model units and μ s.

3 Results

3.1 Normal hearing (NH) acoustic model

3.1.1 Pure tones

Figure 5 shows a direct comparison of the model predictions and behavioural data. Note that the scales are measures of different quantities, specifically for human behavioural responses the ITD threshold is given in μs , whereas the model has arbitrary model units. We observe that the high frequency limit for ITD detection reported from Brughera et al. (2013) is consistent with the model prediction of JND in ITD. However, we also observe that the low frequency ITD detection threshold is estimated by the model to be lower than what the behavioural data indicates. The correlation between the model prediction and the behavioural data of $r = 0.91$ ($p < 0.01$) indicates a good correspondence between the model predictions and behavioural data.

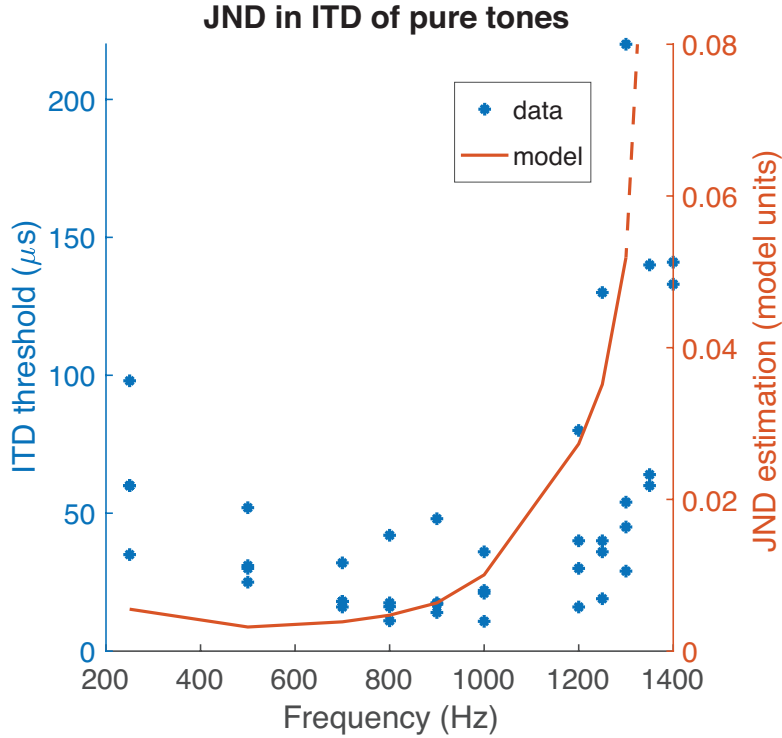


Figure 5: Pure tone model performance. Experimental data from Brughera et al. (2013). The stimulus duration was 300 ms at an intensity of 75 dB SPL. Note that the term *model units* used here and throughout the manuscript represents a dimensionless measure of JND, which is consistent across behavioural data for various stimuli. The JND estimation is inaccurate above 1300 Hz (where the high frequency limit exists), which is shown in dotted lines.

3.1.2 Sinusoidally amplitude modulated (SAM) tones

Figure 6 shows a direct comparison between behavioural and model data. The model correctly predicts the trend of human behavioural responses for variation in the modulation frequency of SAM tones (Bernstein and Trahiotis, 2002). Specifically, ITD thresholds go up for low and high f_m and they reach a minimum in the region of 100–200 Hz. Note

that for all f_c the high frequency cut-off is predicted to be not as steep as behavioural data indicate. Moreover, specifically for $f_c = 4$ kHz, there is some extent of f_m mismatch between the minimum ITD threshold for human behavioural responses and the minimum ITD threshold estimated by the model. Further to visual inspection, this frequency mismatch was identified with a change in the correlation value. If we maintain equivalence between the modulation frequency axis between model and data we get a measure of $r = -0.2837$. However when we translate the model outcome along the f_m axis so as to align the minima of model and data we get a correlation of $r = 0.89$ ($p = 0.11$). For $f_c = 6$ kHz, we observe a good match without any frequency mismatch, which is also indicated by a high correlation value ($r = 0.89$, $p = 0.04$). For $f_c = 10$ kHz, we observe a good match for low modulation frequencies (≤ 125 Hz) whereas for higher modulation frequencies the model predicts higher sensitivity (i.e., lower ITD threshold values). We also observe that when changing f_c the model does not predict the reported behavioural data (Bernstein and Trahiotis, 2002). Behaviourally, an increase of f_c increases ITD thresholds. The model is unable to capture this change. This shortcoming is also reflected in the lower correlation $r = 0.58$ ($p = 0.04$) between behavioural data and model predictions across all f_c .

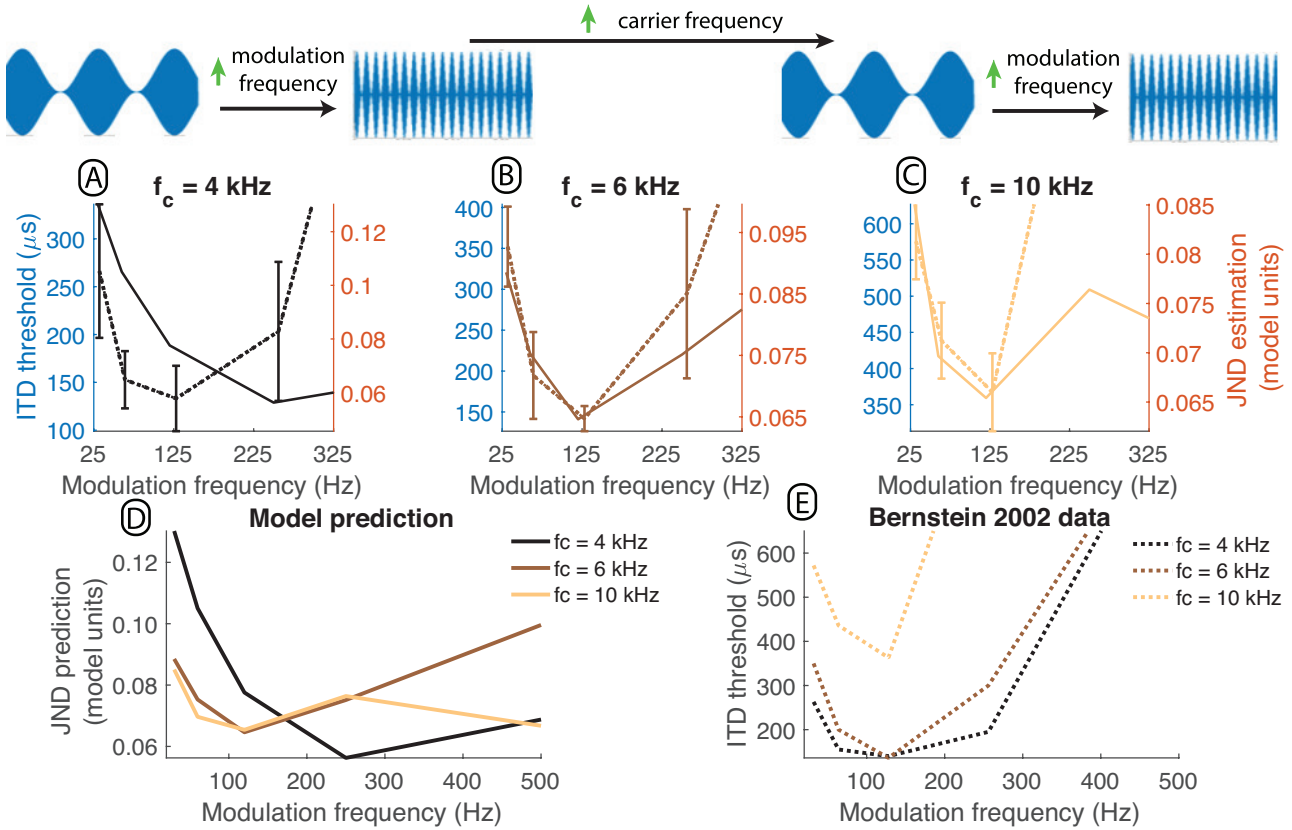


Figure 6: SAM tones model performance. Experimental data from Bernstein and Trahiotis (2002). The stimulus duration was 300 ms at an intensity of 75 dB SPL. Panels A, B, and C compare behavioural data (dotted lines) with model predictions (solid lines) for SAM tones with $f_c = 4, 6, 10$ kHz (respectively). The change in f_c essentially translates the region of neuronal stimulation on the basilar membrane. Panel D and E show the model predictions and behavioural data, respectively, across different f_c .

3.1.3 Transposed tones

Figure 7 shows behavioural and model data. Comparably to SAM tones, we observe again a similar trend for both the model and the behavioural data where the ITD threshold is minimised for a particular f_m region. This f_m region is predicted to be lower than the behavioural data (Bernstein and Trahiotis, 2002). The overall trend for individual carrier frequencies is well predicted with correlation values of $r = 0.81$ ($p = 0.09$), $r = 0.87$ ($p = 0.05$) and $r = 0.93$ ($p = 0.02$) for centre frequencies of 4 kHz, 6 kHz and 10 kHz respectively. However, similarly to the SAM tone case, the across centre frequency variation is not well predicted with a correlation of $r = 0.55$ ($p = 0.07$).

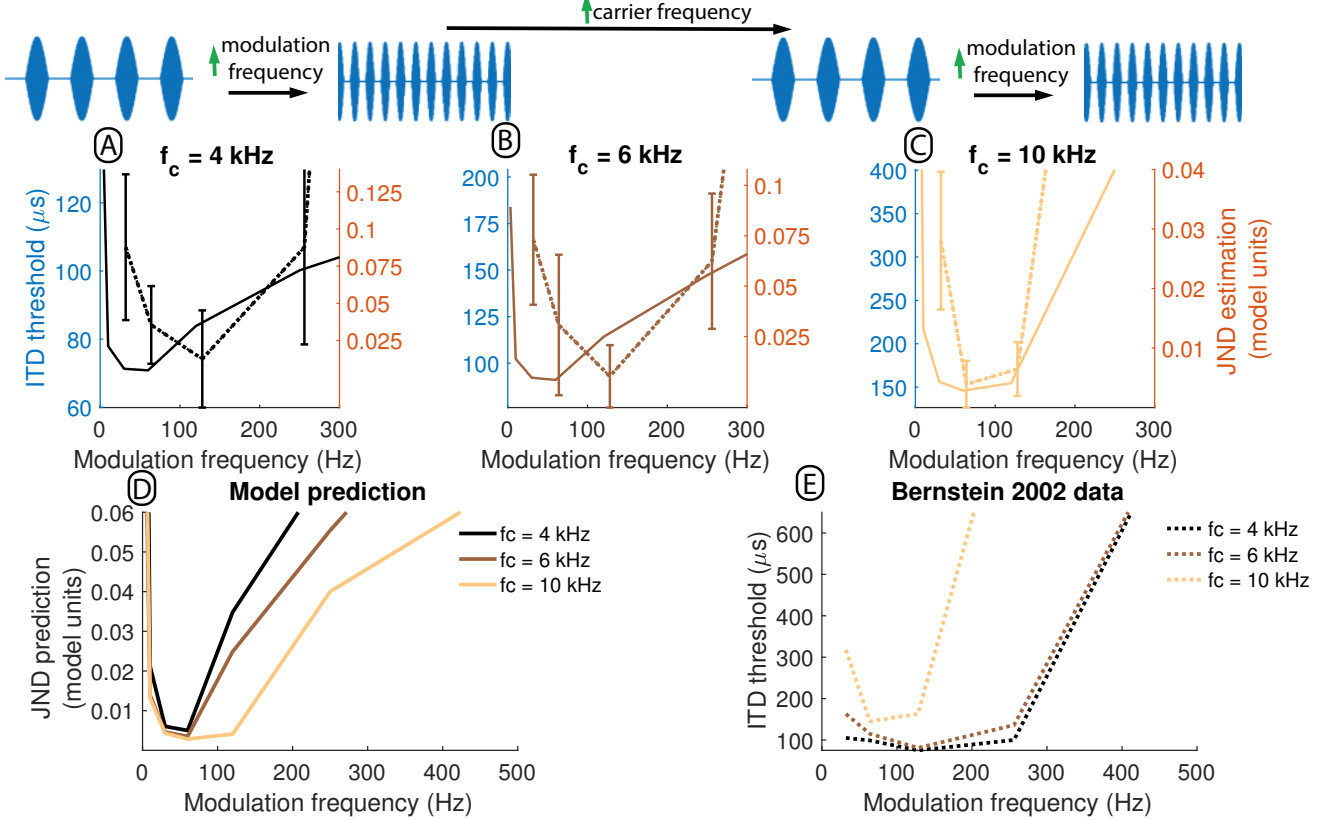


Figure 7: Transposed tones model performance. Experimental data from Bernstein and Trahiotis (2002). The stimulus duration was 300 ms at an intensity of 75 dB SPL. Panels A, B, and C compare behavioural data (dotted lines) with model predictions (solid lines) for SAM tones with $f_c = 4, 6, 10$ kHz (respectively). Panel D and E show the model predictions and behavioural data, respectively, across different f_c .

The modelling outcomes indicate that the ITD threshold is lower for transposed stimuli compared with SAM tones, which agrees with reported observations (Bernstein and Trahiotis, 2002). However, the model predicts a larger relative difference between transposed tones and SAM tones, (Fig. 8).

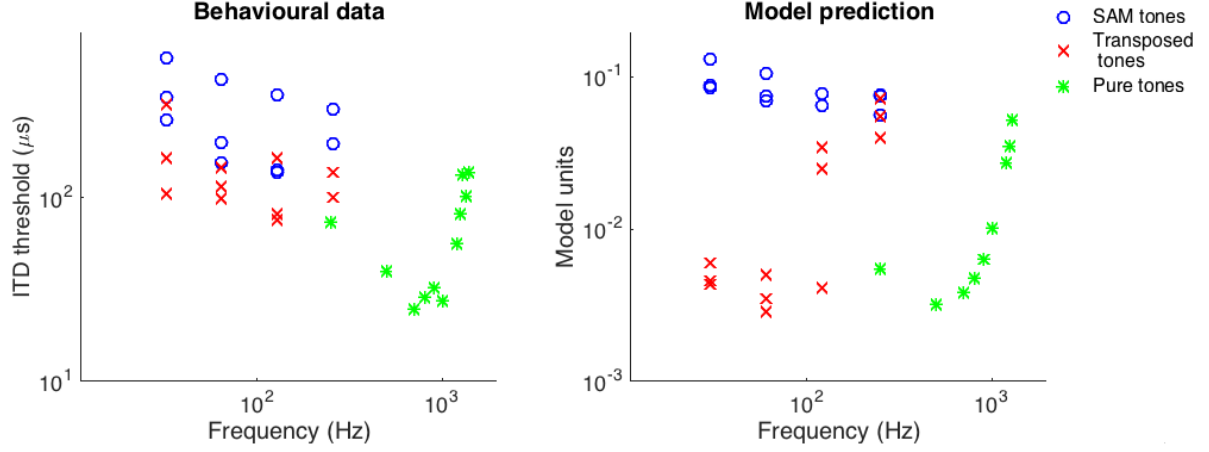


Figure 8: Scatter plot of the mean of behavioural data (left) and model prediction (right) for SAM tones (Bernstein and Trahiotis, 2002) and transposed tones (Bernstein and Trahiotis, 2002) at various modulation frequencies, and pure tones (Brughera et al., 2013) at various centre frequencies. The abscissa is either pure tone frequency or modulation frequency in the case of SAM and transposed tones.

3.1.4 Raised sine modulation

Figure 9 shows a direct comparison of model prediction and behavioural data. Behavioural measures indicate a decrease in ITD thresholds as the modulation exponent is increased for various modulation frequencies (Bernstein and Trahiotis, 2009), which is predicted by the model with a correlation of $r = 0.60$ ($p = 0.01$), across exponent values. Similarly to the previous modulated tones the behavioural data indicate a certain modulation frequency region where the ITD threshold is minimized. However, similarly to the transposed tones, the frequency region which minimises ITD thresholds is predicted by the model to be lower than the behavioural data reported (Bernstein and Trahiotis, 2009). If we account for this frequency mismatch by shifting the model prediction along the f_m axis then the correlation values for the individual exponent values are $r = 0.88$ ($p = 0.12$), $r = 0.98$ ($p = 0.02$), $r = 0.77$ ($p = 0.22$) and $r = 0.98$ ($p < 0.01$) for exponent equal to 1, 2, 4 and 8 respectively, indicating a good overall trend prediction.

Additionally, the effect of decreasing the modulation depth was investigated, as shown in Fig. 10. The behavioural data show a reduction in ITD threshold as the modulation depth increases, which is also indicated by the model, with correlation values of $r = 0.89$ ($p = 0.11$), $r = 0.94$ ($p = 0.06$), and $r = 0.57$ ($p = 0.43$) for exponents 1, 1.5 and 8 respectively. However as the exponent increases for the various modulation depths, the model predicts that the ITD threshold increases, which is contrary to the reported data (Bernstein and Trahiotis, 2009), and is reflected in a low correlation value of $r = 0.46$ ($p = 0.13$).

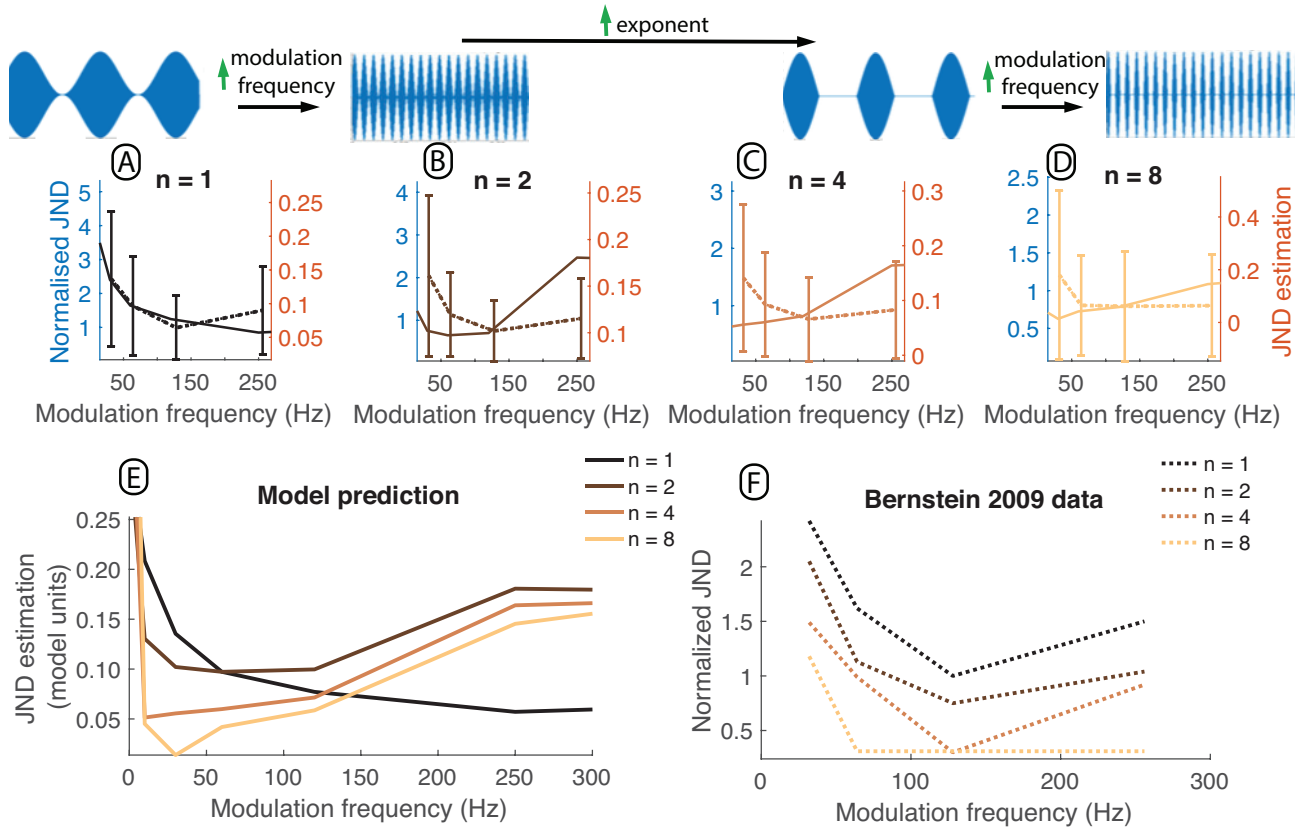


Figure 9: Model performance for raised sine with variable modulation frequency. Behavioural data from Bernstein and Trahiotis (2009). The stimulus duration was 300 ms at an intensity of 75 dB SPL. The normalization in the behavioural data was accomplished by dividing an individual listener's JND by the same listener's JND at 128 Hz modulated SAM tone. This was done so as to remove inter-subject variability (Bernstein and Trahiotis, 2009). Panels A, B, C, and D compare behavioural data (dotted lines) with model predictions (solid lines) for raised sine tones with an exponent equal to 1, 2, 4, and 8 (respectively). Panel E and F show the model predictions and behavioural data, respectively, across different exponent values.

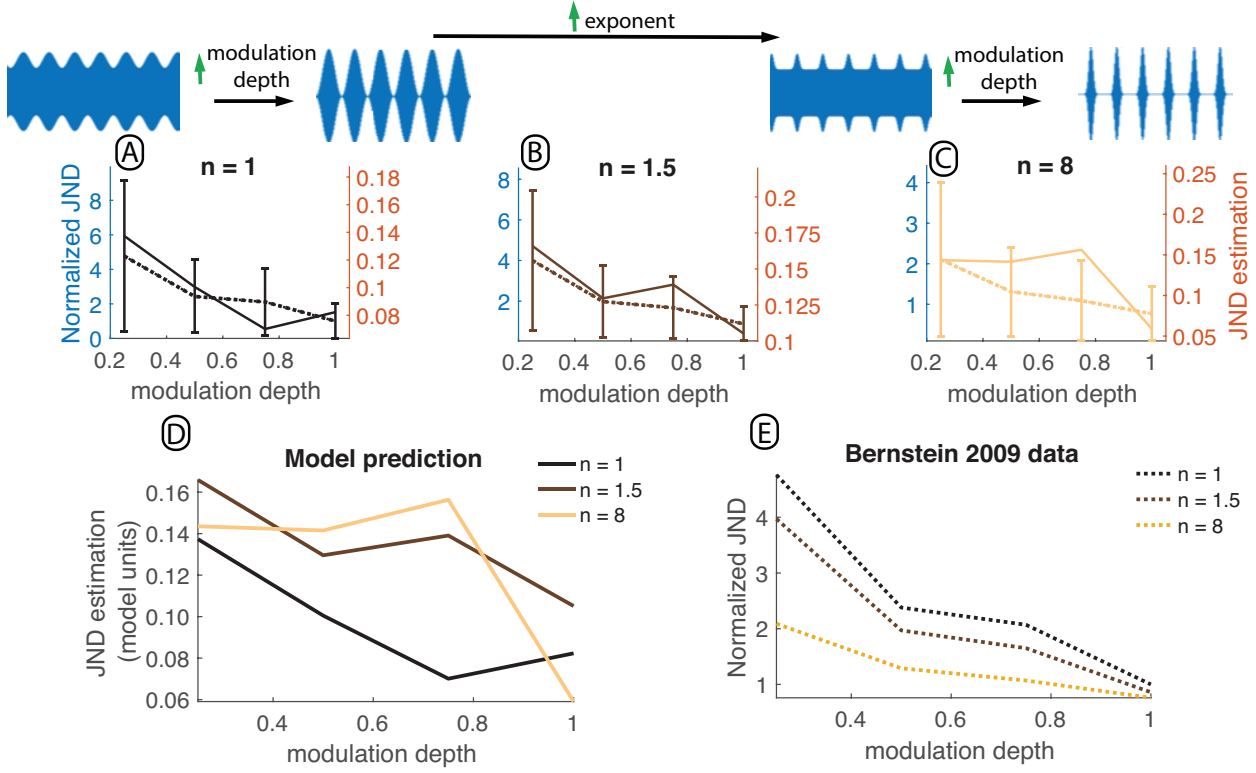


Figure 10: Model performance for raised sine with variable modulation depth. Experimental data from Bernstein and Trahiotis (2009). The stimulus duration was 300 ms at an intensity of 75 dB SPL. The normalization of the behavioural data is done as described in the caption of Fig. 9. Panels A, B, and C compare behavioural data (dotted lines) with model predictions (solid lines) for raised sine tones with an exponent equal to 1, 2, and 8 (respectively). Panel D and E show the model predictions and behavioural data, respectively, across different exponent values.

3.1.5 Trapezoidal modulation

Figure 11 illustrates a direct comparison between model predictions and behavioural data. The behavioural data reported indicate that by increasing the ‘off time’ the ITD detection thresholds decrease (Laback et al., 2011). The model output correctly indicates the same trend as observed in the behavioural data, specifically the correlation value between model prediction and behavioural data is $r = 0.97$ ($p = 0.03$), $r = 0.95$ ($p = 0.01$), $r = 0.93$ ($p = 0.02$), and $r = 0.94$ ($p = 0.02$) when the slope is 6, 8, 10 and 12 dBms⁻¹ respectively. A further observation was that by increasing the rising and falling slope the ITD detection thresholds decrease (Laback et al., 2011). When comparing across all the slope conditions the correlation value is $r = 0.82$ ($p < 0.01$), which indicates good prediction from the model, albeit a higher sensitivity to slope changes is indicated by the model than the behavioural data.

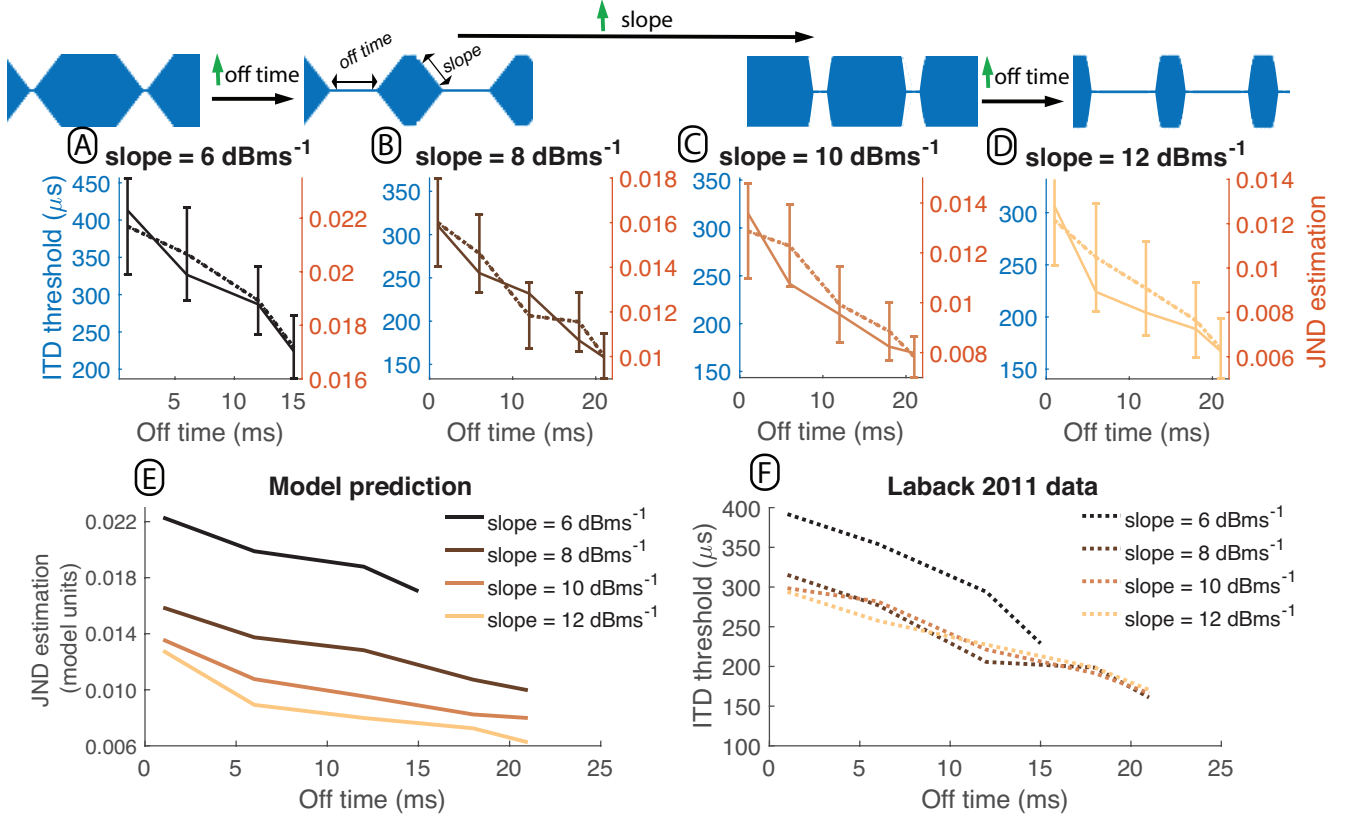


Figure 11: Model performance for trapezoid modulation. Experimental data from Laback et al. (2011). The stimulus duration was 1 s at an intensity of 75 dB SPL. Panels A, B, C, and D compare behavioural data (dotted lines) with model predictions (solid lines) for trapezoid modulated tones with with an onset and offset slope of 6, 8, 10, and 12 dB ms⁻¹ (respectively). Panel E and F show the model predictions and behavioural data, respectively, across the various onset/offset slope values.

3.2 Electric hearing model validation

3.2.1 Unmodulated low frequency pulse trains

The stimulation rate in PPS was varied to test the ITD perception as a function of pulse rate. Figure 12 shows a comparison of model prediction against behavioural data.

The behavioural data indicate that for rates higher than 100 PPS, the majority of BiCI users start having larger ITD detection thresholds. which become unmeasurable for rates higher than 400–800 PPS (Laback et al., 2015). The model correctly predicts that for a certain region of low rate stimulation the ITD detection threshold is constant and it starts increasing after a certain point and becomes physiologically non-relevant for rates higher than 800 PPS. However, it estimates lower ITD detection thresholds than the behavioural data in the region of 400–800 PPS. Overall, the model predicts well the behavioural trend with a correlation value of $r = 0.91$ ($p = 0.002$) between model predictions and behavioural data.

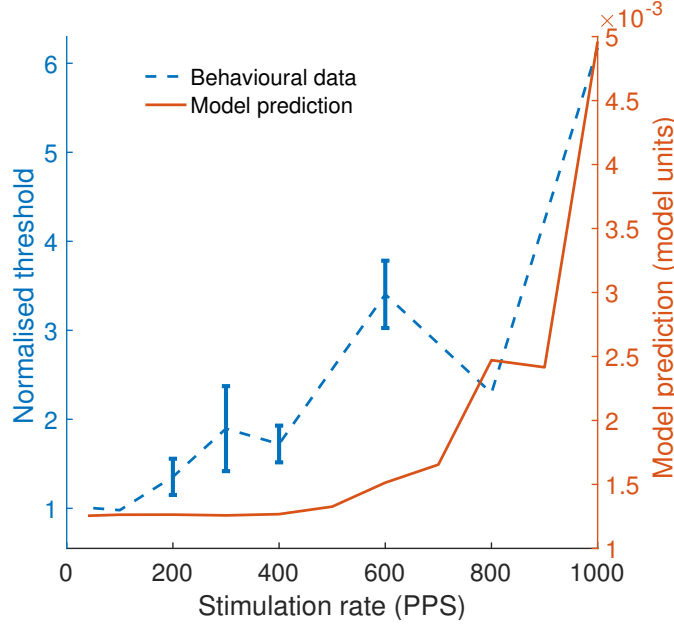


Figure 12: Model performance with unmodulated electric pulses. Behavioural data were extracted from six experiments (Egger et al., 2016; Laback et al., 2007; van Hoesel and Clark, 1997; van Hoesel and Hoesel, 2007; Van Hoesel et al., 2009; van Hoesel and Tyler, 2003) summarised in Fig. 3 of Laback et al. (2015). The median for each study is estimated across all listeners per study and the resulting means normalised across all studies. The normalised medians are further averaged into a single mean which is shown along with the error bars that indicate the standard error of the mean across the six experiments. This is done to remove any ITD threshold offsets that are imposed by differences in the experimental procedure and does not affect variations due to the stimulation rate. The stimulus duration was 300 ms and the maximum electric current was set to drive the neurons at a firing efficiency of 1.25 dB (which is around 75% chance to generate an action potential for a single maximum current pulse).

3.2.2 Sinusoidally amplitude modulated (SAM) pulse trains

Figure 13 shows the comparison of behavioural data with the model prediction. The behavioural result reported was a v-shaped curve of ITD threshold versus f_m (Noel and Eddington, 2013). This is somewhat predicted by the model with the presence of the notch, however the notch has a different shape and there is a mismatch of the frequency at the minimum point. The behavioural data indicate an ITD threshold minimum at around 100 Hz, while the model predicts such minimum at a higher frequency, around 300 Hz. Furthermore, the model shows an increase in the ITD threshold when the carrier stimulation rate was increased from 1000 PPS to 5000 PPS (i.e. when $f_m > 16$ Hz) which is not indicated by the behavioural data. The model does not predict well the behavioural data and this is indicated by a low correlation value between model and data of $r = 0.26$ ($p = 0.57$).

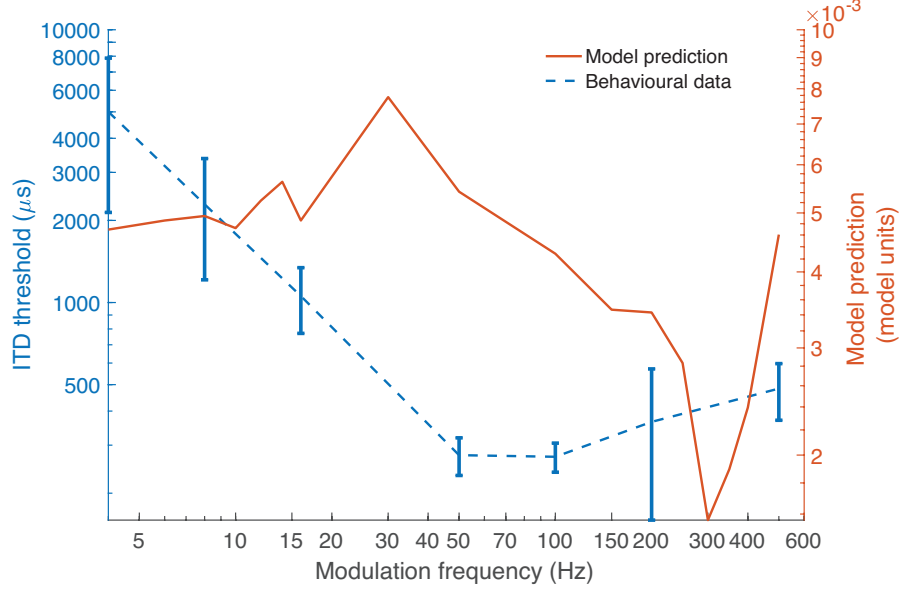


Figure 13: Model performance for electric SAM pulses. Experimental data from (Noel and Eddington, 2013). The stimulus duration was 300 ms and the maximum electric current was set to drive the neurons at a firing efficiency of 1.25 dB, which is around 75% chance of an action potential generated for a single maximum current pulse. Note that for $f_m \leq 16$ Hz the pulse rate was at 1000 PPS, whereas for $f_m > 16$ Hz the pulse rate was at 5000 PPS.

3.2.3 Trapezoidal modulation

Similarly to the acoustic case of the trapezoidal modulation the modulating envelope is a series of symmetric trapezoids defined by ‘off time’, rising and falling slopes and a fixed modulation period of 27.3 Hz. As Fig. 14 shows, the behavioural data indicate that as the ‘off time’ is increased the ITD detection threshold decreases. The model correctly predicts the decrease in ITD threshold as ‘off time’ increases, with correlation values of $r = 0.88$ ($p = 0.12$), $r = 0.90$ ($p = 0.10$) and $r = 0.91$ ($p = 0.09$) for when the slope is 6, 8 and 12 dBms⁻¹ respectively. However, contrary to the acoustic case, the rising and falling slopes do not appear to have significant effect in the ITD detection threshold. This is also well predicted by the model as across the slope correlation has a value of $r = 0.82$ ($p = 0.001$).

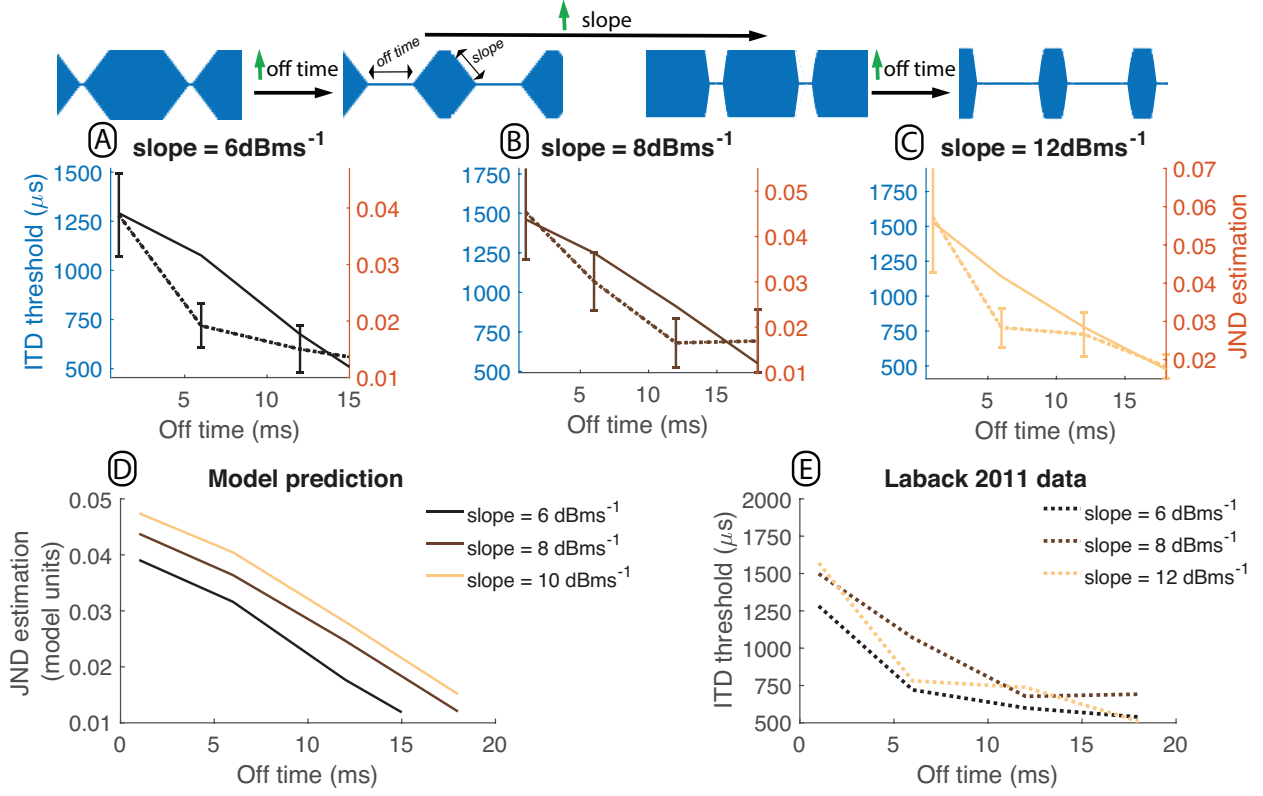


Figure 14: Model performance for electric trapezoid modulation pulses. Experimental data from Laback et al. (2011). The stimulus duration was 1 s and the maximum electric current was set to drive the neurons at a firing efficiency of 1.25 dB, which is around 75% chance of an action potential generated for a single maximum current pulse (Sec. 2.1.2). Panels A, B, C, and D compare behavioural data (dotted lines) with model predictions (solid lines) for trapezoid modulated tones with with an onset and offset slope of 6, 8, 10, and 12 dB ms⁻¹ (respectively). Panel E and F show the model predictions and behavioural data, respectively, across the various onset/offset slope values.

4 Discussion

Model purpose and assumptions

We combined acoustic and electric modelling that describe the temporal properties of action potential generation on the AN from both acoustic and electric stimulation with a novel neurometric-psychometric method. Its purpose is to map the temporal pattern of action potentials to the perceptual sensation of ITD (Stüttgen et al., 2011) by estimating the JND in ITD. The binaural ITD sensitivity modelling approach taken here is functional/phenomenological and, as such, a qualitative comparison with behavioural data is feasible. The selection of the various studies providing the behavioural data creates an evaluation framework for the model.

The model presented here is comparable with other models in proposing a generic binaural processor (Breebaart et al., 2001; Dietz et al., 2011; Pulkki and Hirvonen, 2009; Takanen et al., 2014). These models typically consider more complex stimuli with variable ILD and broadband spectrum. However, we wanted to investigate simpler stimuli to reduce the assumptions made about how the central auditory system processes binaural information since there is no paradigm yet that explicitly links human ITD detection per-

formance and ascending neural pathways in the central auditory system. Furthermore, previous correlation-based approaches investigate signals which are continuous representations of periphery (Colburn, 1973; Colburn and Durlach, 1978), whereas here we investigate a discrete representation consisting of binary events. Because a direct application of the short-time interaural cross-correlation (Colburn, 1973) is not applicable, we used the SCC (Joris et al., 2006). This binary events analysis enables the direct comparison of spike trains generated on the AN regardless of the nature of the stimulus, either it being acoustic or electric.

Neurons typically have intrinsic noise in their behaviour and, as such, the model output has a degree of variance. In order to minimize this variance of the model output we assumed around 1000 neurons with similar temporal characteristics originating from the periphery and converging to one processing centre which is perfectly binaurally place matched. This assumption could be biophysically unrealistic, especially when considering electrical stimulation, as CI subjects frequently have incomplete neural survival. A further consideration for contrasting the BiCI results and the model prediction is intrinsic variation in behavioural data. This is due to the low number of subjects in the behavioural data, often from a single study. Furthermore, there is typically much unaccounted variability between CI subjects, either due to pathology of the AN and the sound deprived brain, variation in place/locus of stimulating electrodes, or plasticity induced post-operative changes of the tonotopic map.

Model performance assessment and future steps

The model performance was tested by comparing its output against behavioural data. For the NH situation, the model predicted the general trends observed for all waveforms tested.

However, the model was found to be more sensitive to temporal gaps in stimulus (i.e. periods of ‘off time’) than the behavioural data indicates. Specifically, this is observed from transposed tones and raised sine tones, where by varying the f_m the model consistently predicts that the minimum f_m for ITD sensitivity is at a lower frequency (i.e., larger temporal gap) than what the behavioural data indicates. This might be caused by the peripheral model’s high entrainment to temporal gaps of high intensity stimuli. Further investigations on internal noise, possibly by modelling central processes, can reduce this high entrainment the model exhibits.

The current NH peripheral model uses neurons of a single type of spontaneous activity whereas in reality the AN has a mixture of high spontaneous rate, medium spontaneous rate and low spontaneous rate neurons (Relkin and Doucet, 1991; Rhode and Smith, 1985). The single neuron type was chosen for the sake of simplicity. A preliminary analysis of low spontaneous rate neurons indicated reduced response of neurons to the onset of modulating periods. This affects the model outcome by estimating higher ITD JND values. Model outcomes when using mixed type neurons, specifically 60% high spontaneous rate, 20% medium spontaneous rate and 20% low spontaneous rate, showed little variation from the reported results which used 100% high spontaneous rate neurons (see Appendix). However, this might be attributed to the high intensity stimuli used in the behavioural studies that form the evaluation framework of the model. Low intensity stimuli could be used to better describe the possible role of low spontaneous rate neurons in ITD detection.

Moreover, the model does not predict the behavioural changes in ITD perception across CF (i.e., different locations on the basilar membrane) reported for SAM and trans-

posed tones. This can be attributed to the fact that no tonotopic considerations were made for the binaural processing of more central auditory pathways. A recent biophysical model which investigated ITD sensitivity was proposed by Chung et al. (2015). They found that their model of brainstem neurons required stronger excitatory synaptic inputs and faster membrane responses for ITD sensitivity at high stimulation rates, whereas weaker excitatory synaptic input and slower membrane responses were necessary at low stimulation rates, for both electric and acoustic stimulation. Their findings suggest the possibility of frequency-dependent differences in the neural mechanisms of binaural processing, which could support the need for a filter, especially for the spectral aspects (i.e., introduce tonotopic considerations) of the proposed model.

For the BiCI situation, the model predicts well the ITD sensitivity for unmodulated pulse trains and temporal variations of trapezoid modulations. However, it does not predict well SAM pulse train ITD sensitivity. This mismatch between model prediction and behavioural data could be attributed to the electric stimulation peripheral model, which is unable to predict the exact variability in action potential generation as a function of pulse rate and firing rate. It is also unable to quantitatively predict sensitivity of AN fibres to small modulation depths (Goldwyn et al., 2012).

A consideration that needs to be made when examining behavioural data for BiCI, especially for SAM pulse trains, is that more data are required, since the data presented are from a single study with 5 subjects (Noel and Eddington, 2013).

Caution needs to be taken when interpreting the correlation values between model and data, especially for the cases where a single condition is examined (e.g. single centre frequency, single trapezoid slope, etc.) because of the low number of data points, typically restricted to 4-6 measurements of variable changes. Across conditions, the number of data points increases to 12-16, based on the number of conditions that were tested.

Further improvement is necessary for the electric stimulation peripheral model. The complete characterization of neuronal responses to electrical stimulation is ongoing (e.g., Boulet et al., 2016). However, this does not directly affect the formulation of the proposed model. Any peripheral model which can produce a sequence of binary events describing the spatiotemporal pattern (i.e., answering the ‘where?’ and the ‘when?’) of AN activity could be used in the pipeline of the proposed framework. A more encompassing electric peripheral model is a part of future improvements to this model.

5 Conclusion

In summary, the trend of variations of JND in ITD caused by frequency changes in *unmodulated* stimuli, (i.e., pure tones and electric pulse trains of clinical stimulation rates) are well described by the model. Specifically, the high frequency cut-off for ITD detection for both acoustic and electric hearing is predicted. Therefore the model can identify the useful region of ITD detection for these types of stimuli. Furthermore, the trend of variations of JND in ITD caused by *modulated* stimuli with low frequency modulation (i.e., ≤ 30 Hz) is also well described. For mid-high modulation frequencies (i.e., ≥ 100 Hz) the model indicates the general trends observed. However it underestimates the behavioural data (i.e. predicts lower JNDs), hence its main strength lies in qualitatively predicting the effect of a change in stimulus parameter on ITD detection, rather than quantitative prediction of JND in ITD across all possible stimuli.

The applicability of the model proposed here on both BiCI and NH behavioural data support the working hypothesis of a normal operation of the central system for binaural

detection, which is also made by Chung et al. (2015). However it should be noted that extended periods of deafness could invalidate this hypothesis. This could cause the general poor ITD sensitivity observed for subjects with extended periods of binaural sensory deprivation (Litovsky et al., 2010).

The model presented here is only suitable for processing narrowband stimuli (i.e., within an ERB). Equivalently for the electrical stimulation case, only a single electrode interaural pair can be considered as there is no explicit modelling of cross-channel interactions. Furthermore, no ILDs were considered for this analysis. These limitations pose no problem when studying static ITD perception of narrowband stimuli, on the contrary they enable fair comparison between model predictions and behavioural data. As such, the psychoacoustic data used in the evaluation framework for model validation used static ITD with narrowband stimuli, or single electrode pair for the electric case, and minimised ILD cues.

A computational model aids exploring the implications of modifying the stimulus waveform on binaural ITD detection. As such it can be used as a metric for optimization of signal processing schemes in auditory prostheses. A stimulus waveform which emphasises the coding of different temporal cues in running speech (such as fundamental frequency, speech envelope, onsets in speech, e.g. Francart et al. (2014) has the potential to improve sound localization and speech perception in noise, by providing the BiCI user with better ITD cues alongside ILD cues.

Acknowledgements

We thank Ville Pulkki, Torsten Dau, Bastian Epp and Suyash Joshi for stimulating discussions. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme FP7/2007-2013/ under REA grant agreement PITN-GA-2012-317521 “ICanHear”. Authors AP and AMT are fellows. Support was also provided by a research grant (150432 “Advances in auditory implants: signal processing and clinical aspects”) by the Flemish Agency for Innovation by Science and Technology (IWT).

References

- Bernstein, L. R. and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies by using “transposed stimuli”. *The Journal of the Acoustical Society of America*, 112(3).
- Bernstein, L. R. and Trahiotis, C. (2009). How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli. *The Journal of the Acoustical Society of America*, 125(5).
- Boulet, J., White, M., and Bruce, I. C. (2016). Temporal considerations for stimulating spiral ganglion neurons with cochlear implants. *Journal of the Association for Research in Otolaryngology*, 17(1):1–17.
- Breebaart, J., van de Par, S., and Kohlrausch, A. (2001). Binaural processing model based on contralateral inhibition. i. model structure. *The Journal of the Acoustical Society of America*, 110(2).

- Briaire, J. J. and Frijns, J. H. (2000). Field patterns in a 3D tapered spiral model of the electrically stimulated cochlea. *Hearing Research*, 148(1–2):18 – 30.
- Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, 86(1):117–128.
- Bronkhorst, A. W. and Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *Acoustical Society of America Journal*, 83:1508–1516.
- Bruce, I. C., Irlicht, L. S., White, M. W., O’Leary, S. J., Dynes, S., Javel, E., and Clark, G. M. (1999). A stochastic model of the electrically stimulated auditory nerve: pulse-train response. *IEEE Transactions on Biomedical Engineering*, 46(6):630–637.
- Brughera, A., Dunai, L., and Hartmann, W. M. (2013). Human interaural time difference thresholds for sine tones: The high-frequency limit. *The Journal of the Acoustical Society of America*, 133(5):2839–2855.
- Cartee, L. A., Miller, C. A., and van den Honert, C. (2006). Spiral ganglion cell site of excitation I: Comparison of scala tympani and intrameatal electrode responses. *Hearing Research*, 215(1–2):10–21.
- Chung, Y., Delgutte, B., and Colburn, H. S. (2015). Modeling binaural responses in the auditory brainstem to electric stimulation of the auditory nerve. *Journal of the Association for Research in Otolaryngology*, 16(1):135–158.
- Colburn, H. S. (1973). Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. *The Journal of the Acoustical Society of America*, 54(6):1458–1470.
- Colburn, H. S. and Durlach, N. I. (1978). Models of binaural interaction. *Handbook of perception*.
- Dietz, M., Ewert, S. D., and Hohmann, V. (2011). Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication*, 53(5):592–605.
- Egger, K., Majdak, P., and Laback, B. (2016). Channel Interaction and Current Level Affect Across-Electrode Integration of Interaural Time Differences in Bilateral Cochlear-Implant Listeners. *JARO - Journal of the Association for Research in Otolaryngology*, 17(1):55–67.
- Faisal, A. A., Selen, L. P. J., and Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, 9(april):292–303.
- Firszt, J. B., Reeder, R. M., and Skinner, M. W. (2008). Restoring hearing symmetry with two cochlear implants or one cochlear implant and a contralateral hearing aid. *Journal of rehabilitation research and development*, 45(5):749–767.
- Francart, T., Brokx, J., and Wouters, J. (2009). Sensitivity to interaural time differences with combined cochlear implant and acoustic stimulation. *JARO - Journal of the Association for Research in Otolaryngology*, 10(1):131–141.

- Francart, T., Lenssen, A., and Wouters, J. (2011). Sensitivity of bimodal listeners to interaural time differences with modulated single- and multiple-channel stimuli. *Audiology and Neurotology*, 16(2):82–92.
- Francart, T., Lenssen, A., and Wouters, J. (2014). Modulation enhancement in the electrical signal improves perception of interaural time differences with bimodal stimulation. *JARO - Journal of the Association for Research in Otolaryngology*, 15(4):633–647.
- Glasberg, B. R. and Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1-2):103–138.
- Goldwyn, J. H., Rubinstein, J. T., and Shea-Brown, E. (2012). A point process framework for modeling electrical stimulation of the auditory nerve. *Journal of Neurophysiology*, 108(5):1430–1452.
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species - 29 years later. 87(6):2592–2605.
- Joris, P. X. (2003). Interaural time sensitivity dominated by cochlea-induced envelope patterns. *The Journal of Neuroscience*, 23(15):6345–6350.
- Joris, P. X., Louage, D. H., Cardoen, L., and van der Heijden, M. (2006). Correlation Index: A new metric to quantify temporal coding. *Hearing Research*, 216-217:19–30.
- Kidd, G., Mason, C. R., Best, V., and Marrone, N. (2010). Stimulus factors influencing spatial release from speech-on-speech masking. *Acoustical Society of America*, 128(4):1965–1978.
- Laback, B., Egger, K., and Majdak, P. (2015). Perception and coding of interaural time differences with bilateral cochlear implants. *Hearing Research*, 322:138–150.
- Laback, B., Majdak, P., and Baumgartner, W.-D. (2007). Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing. *J Acoust Soc Am*, 121(4):2182.
- Laback, B., Zimmermann, I., Majdak, P., Baumgartner, W.-D., and Pok, S.-M. (2011). Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearing. *J Acoust Soc Am*, 130(3):1515–1529.
- Lieberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber.
- Litovsky, R. Y., Jones, G. L., Agrawal, S., and van Hoesel, R. (2010). Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans. *J Acoust Soc Am*, 127(1):400–414.
- Louage, D. H. G., van der Heijden, M., and Joris, P. X. (2004). Temporal properties of responses to broadband noise in the auditory nerve. *Journal of neurophysiology*, 91(5):2051–2065.
- Macpherson, E. A. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J Acoust Soc Am*, 111(5 Pt 1):2219–2236.

- McNeal, D. R. (1976). Analysis of a Model for Excitation of Myelinated Nerve. *IEEE Transactions on Biomedical Engineering*, BME-23(4):329–337.
- Miller, C. A., Robinson, B. K., Rubinstein, J. T., Abbas, P. J., and Runge-Samuelson, C. L. (2001). Auditory nerve responses to monophasic and biphasic electric stimuli. *Hearing Research*, 151(1-2):79–94.
- Mino, H., Rubinstein, J. T., Miller, C. A., and Abbas, P. J. (2004). Effects of Electrode-to-Fiber Distance on Temporal Neural Response with Electrical Stimulation. *IEEE Transactions on Biomedical Engineering*, 51(1):13–20.
- Mino, H., Rubinstein, J. T., and White, J. A. (2002). Comparison of algorithms for the simulation of action potentials with stochastic sodium channels. *Annals of Biomedical Engineering*, 30(4):578–587.
- Motz, H. and Rattay, F. (1986). A study of the application of the Hodgkin-Huxley and the Frankenhaeuser-Huxley model for electrostimulation of the acoustic nerve. *Neuroscience*, 18(3):699–712.
- Nicoletti, M., Rudnicki, M., and Hemmert, W. (2010). A model of the auditory nerve for acoustic- and electric excitation. *Frontiers in Computational Neuroscience*, 4(0):5.
- Nicoletti, M., Wirts, C., and Hemmert, W. (2013). Modeling Sound Localization with Cochlear Implants. In *The Technology of Binaural Listening*, pages 309–331.
- Noel, V. A. and Eddington, D. K. (2013). Sensitivity of bilateral cochlear implant users to fine-structure and envelope interaural time differences. *The Journal of the Acoustical Society of America*, 133(4):2314–28.
- Offeciers, E., Morera, C., Müller, J., Huarte, A., Shalloo, J., and Cavallé, L. (2005). International consensus on bilateral cochlear implants and bimodal stimulation. *Acta Oto-Laryngologica*, 125(9):918–919. PMID: 16109670.
- Pulkki, V. and Hirvonen, T. (2009). Functional count-comparison model for binaural decoding. *Acta Acustica united with Acustica*, 95(5):883–900.
- Rattay, F. (1986). Analysis of Models for External Stimulation of Axons. *IEEE Transactions on Biomedical Engineering*, BME-33(10):974–977.
- Relkin, E. M. and Doucet, J. R. (1991). Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons. *Hearing Research*, 55(2):215–222.
- Rhode, W. S. and Smith, P. H. (1985). Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers. *Hearing Research*, 18(2):159–168.
- Rothman, J. S. and Manis, P. B. (2003). The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons. *Journal of neurophysiology*, 89(6):3097–3113.
- Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences*, 99(5):3318–3323.

- Spoendlin, H. and Schrott, a. (1988). The spiral ganglion and the innervation of the human organ of Corti. *Acta oto-laryngologica*, 105(5-6):403–410.
- Stüttgen, M., Schwarz, C., and Jäkel, F. (2011). Mapping spikes to sensations. *Frontiers in Neuroscience*, 5:125.
- Takanen, M., Santala, O., and Pulkki, V. (2014). Visualization of functional count-comparison-based binaural auditory model output. *Hearing Research*, 309:147–163.
- Van Deun, L., Van Wieringen, A., Francart, T., Scherf, F., Dhooge, I. J., Deggouj, N., Desloovere, C., Van De Heyning, P. H., Offeciers, F. E., De Raeve, L., and Wouters, J. (2009). Bilateral cochlear implants in children: Binaural unmasking. *Audiology and Neurotology*, 14(4):240–247.
- van Hoesel, R. J. and Clark, G. M. (1997). Psychophysical studies with two binaural cochlear implant subjects. *The Journal of the Acoustical Society of America*, 102(1):495–507.
- Van Hoesel, R. J. M. (2012). Contrasting benefits from contralateral implants and hearing aids in cochlear implant users. 288(1-2):100–113.
- van Hoesel, R. J. M. and Hoesel, R. J. M. V. (2007). Sensitivity to binaural timing in bilateral cochlear implant users. *J Acoust Soc Am*, 121(4):2192.
- Van Hoesel, R. J. M., Jones, G. L., and Litovsky, R. Y. (2009). Interaural time-delay sensitivity in bilateral cochlear implant users: Effects of pulse rate, modulation rate, and place of stimulation. *JARO - Journal of the Association for Research in Otolaryngology*, 10(4):557–567.
- van Hoesel, R. J. M. and Tyler, R. S. (2003). Speech perception, localization, and lateralization with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 113(3):1617–1630.
- Wang, L. and Colburn, H. S. (2012). A modeling study of the responses of the lateral superior olive to ipsilateral sinusoidally amplitude-modulated tones. *JARO - Journal of the Association for Research in Otolaryngology*, 13(2):249–267.
- Wang, L., Devore, S., Delgutte, B., and Colburn, H. S. (2014). Dual sensitivity of inferior colliculus neurons to ITD in the envelopes of high-frequency sounds: experimental and modeling study. *Journal of neurophysiology*, 111(1):164–81.
- Wightman, F. L. and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–61.
- Woo, J., Miller, C. A., and Abbas, P. J. (2010). The dependence of auditory nerve rate adaptation on electric stimulus parameters, electrode position, and fiber diameter: A computer model study. *JARO - Journal of the Association for Research in Otolaryngology*, 11(2):283–296.
- Zilany, M. S. a., Bruce, I. C., and Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America*, 135(1):283–286.

Zilany, M. S. A., Bruce, I. C., Nelson, P. C., and Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *The Journal of the Acoustical Society of America*, 126(5):2390.

Zilany, M. S. a. and Carney, L. H. (2010). Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(31):10380–90.

6 Appendix

6.1 Peripheral model

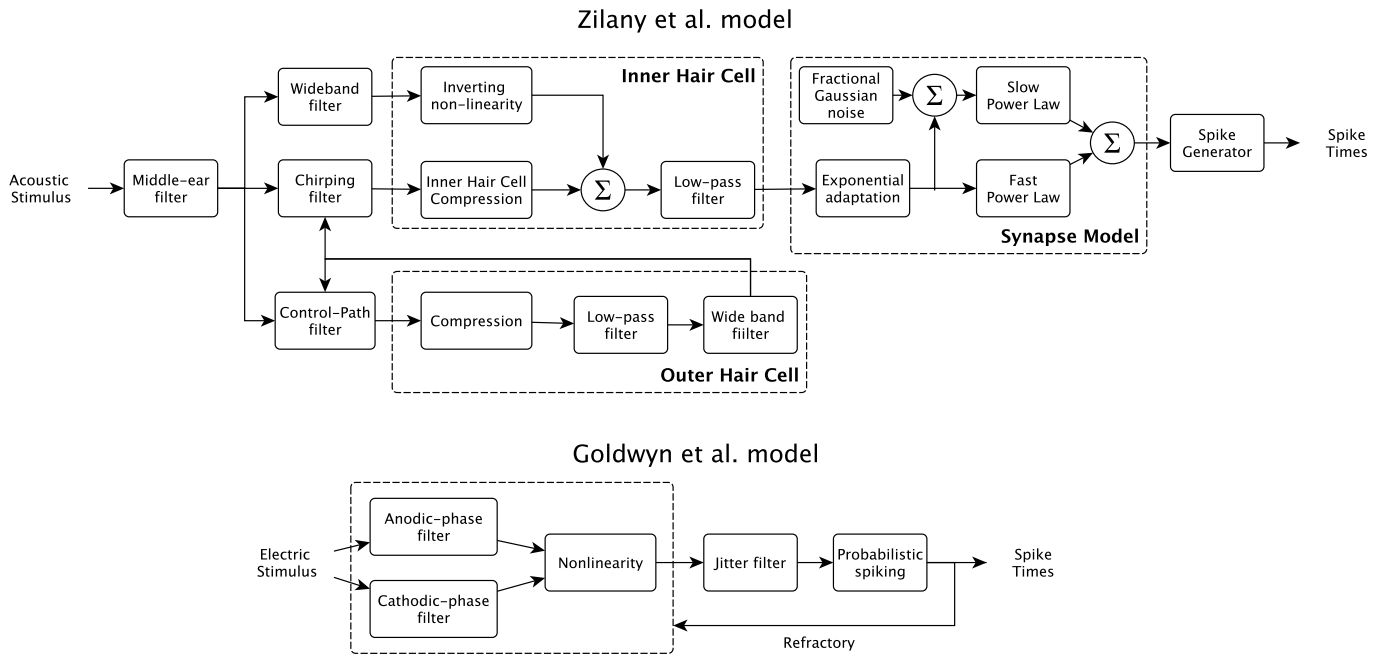


Figure 15: Diagrams of the acoustical and electrical models proposed by Zilany et al. (2014) and Goldwyn et al. (2012), respectively, used as a front end in the proposed framework. The top panel shows the former, which receives an instantaneous pressure waveform (in Pa) as an input. The bottom panel shows the latter, which receives as an input a train of biphasic pulses. Notice that the incoming biphasic pulse train is separated into anodic and cathodic pulses, filtered separately, and then recombined. Both models output AN spike times.

6.2 Stimuli formulation

Pure tones

Pure tones were generated according to Eq. (6), where t is stimulation time and f is the pure tone frequency.

$$\begin{aligned} acou_{left}(t, f, \delta_{left}) &= \sin(2\pi f(t + \delta_{left})) \\ acou_{right}(t, f, \delta_{right}) &= \sin(2\pi f(t + \delta_{right})) \end{aligned} \quad (6)$$

Sinusoidally amplitude modulated (SAM) tones

SAM tone stimuli were generated according to Eq. (7), where t is stimulation time, f_c is the carrier frequency, and f_m is the modulation frequency.

$$\begin{aligned} acou_{left}(t, f_c, f_m, \delta_{left}) &= \sin(2\pi f_c t) \left(2 \left(\frac{\sin(2\pi f_m(t + \delta_{left}))}{2} \right) - 0.5 \right) + 1 \\ acou_{right}(t, f_c, f_m, \delta_{right}) &= \sin(2\pi f_c t) \left(2 \left(\frac{\sin(2\pi f_m(t + \delta_{right}))}{2} \right) - 0.5 \right) + 1 \end{aligned} \quad (7)$$

Transposed tones

Transposed tone stimuli are similar to SAM tones, except that the modulation waveform is rectified. The transposed tones used were generated according to Eq. (8), where t is stimulation time, f_c is the carrier frequency, and f_m is the modulation frequency.

$$\begin{aligned} acou_{left}(t, f_c, f_m, \delta_{left}) &= \sin(2\pi f_c t) \max(0, 2 \left(\left(1 + \frac{\sin(2\pi f_m(t + \delta_{left}))}{2} \right) - 0.5 \right) + 1) \\ acou_{right}(t, f_c, f_m, \delta_{right}) &= \sin(2\pi f_c t) \max(0, 2 \left(\left(1 + \frac{\sin(2\pi f_m(t + \delta_{right}))}{2} \right) - 0.5 \right) + 1) \end{aligned} \quad (8)$$

Raised sine modulation

The formula used for generating raised sine tone stimuli is shown in Eq. (9), where t is stimulation time, f_c is the carrier frequency (which is set to 4 kHz), f_m is the modulation frequency, n is the order by which the sine is raised to, and m is the modulation depth. It is similar to the SAM tone, with the exception that the sinusoidal modulation waveform is raised to a power. This makes the modulated oscillations steeper, hence introducing off time periods between modulation peaks (as shown in Fig. 9). In other words, raising the sinusoidal modulation to a power creates a transition from a SAM tone to a transposed-like tone (Bernstein and Trahiotis, 2009).

$$\begin{aligned} acou_{left}(t, f_m, m, n, \delta_{left}) &= \sin(2\pi f_c t) \left(2m \left(\left(\frac{\sin(2\pi f_m(t + \delta_{left}))}{2} \right)^n - 0.5 \right) + 1 \right) \\ acou_{right}(t, f_m, m, n, \delta_{right}) &= \sin(2\pi f_c t) \left(2m \left(\left(\frac{\sin(2\pi f_m(t + \delta_{right}))}{2} \right)^n - 0.5 \right) + 1 \right) \end{aligned} \quad (9)$$

Trapezoidal modulation

The trapezoid stimulus is an amplitude modulated sinusoid with f_c of 8727 Hz, and an f_m of 27.3 Hz. The modulation envelope is a series of symmetric trapezoids defined by ‘off time’, rising and falling slopes (Fig. 11). Since the f_m is fixed, the repetition period of the trapezoid is constant. Thus, by increasing the ‘off time’ the ‘on time’ is reduced and vice versa. The ‘on time’ is the duration of the stimulus whilst at the peak amplitude.

The transition from the off level to the on level and vice versa is defined by the rising and falling slope respectively. The imposed ITD is only applied on the envelope, meaning that the high frequency carrier has no inter-aural phase difference and only the modulating trapezoids are phase shifted. For a complete detailed description of how the stimulus was generated refer to Laback et al. (2011).

6.3 Electrical stimulation firing efficiency curve

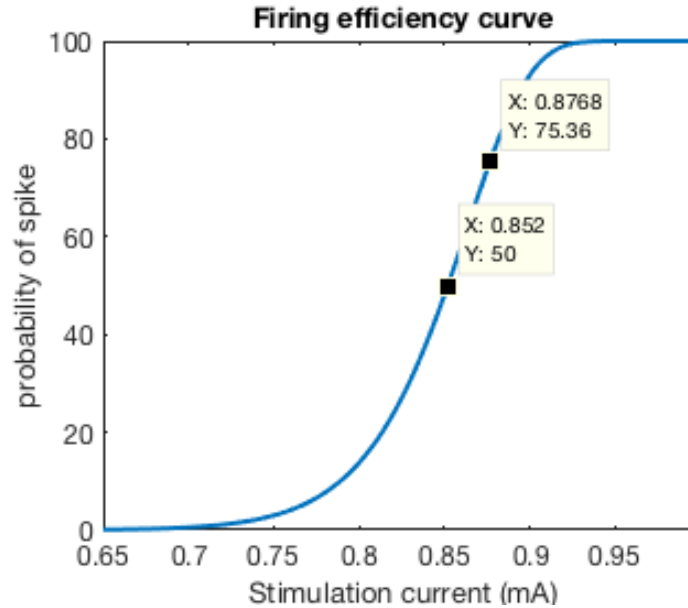


Figure 16: The firing efficiency of the simulated neurons. As shown in the figure above, an input of 0.852 mA has a 50% chance of generating a spike, therefore $I_{thr} = 0.852$ mA. The peak current for each stimulation event, is calculated as $I_{peak} = I_{thr} * 10^{\frac{FE_{dB}-1}{20}}$ (where FE_{dB} is firing efficiency in dB). Therefore, for $FE_{dB} = 1.25$ the value of $I_{peak} = 0.852 * 10^{(1.25-1)/20} = 0.8768$ mA. Note from the firing efficiency curve above that for a stimulation current of 0.8768 mA the probability of generating an action potential is 75.36%.

6.4 Low Spontaneous Rate and High Spontaneous Rate.

Unmodulated stimuli - Pure tones

The Zilany et al. (2014) model exhibits phase locking for high spontaneous rate (HSR) neurons, even at low levels (see Fig. 17 for 75 dB SPL and Fig. 18 for 40 dB SPL). The effect of this phase locking is the clearly defined SCC peaks which identify the periodicity of the signal. Contrary, low spontaneous rate (LSR) neurons exhibit clear phase locking at high stimulus levels (Fig. 17 for 75 dB SPL) but not for low stimulus levels (Fig. 18 for 40 dB SPL).

The BCC method characterises the mismatch of the reference SCC distribution to the target SCC distribution (see Section ‘2.4 Binary classifier characterisation’). The consequence of this is that for the method to give a consistent JND estimate it is necessary for the two SCC distributions to be clearly defined, or in other words it is necessary for the SCC peaks to not be jagged. Such is not the case for the LSR neurons at 40 dB SPL

(Fig. 18), and the outcome of this is that the model fails to predict the JND trend for LSR neurons at 40 dB SPL (see Fig. 19).

The model normalises the SCC peaks so in practice the difference in the firing rate between LSR and HSR is removed. As is shown in Fig. 17 the normalised SCC distributions are quite similar between LSR and HSR, which indicates that the normalised SCC identifies the periodic pattern, and not the firing rate of the auditory nerve. As such, the predominant effect is the overall shape of the SCC distribution which affects the outcome of the model.

What is further observed from Fig. 17 is that the shape of the SCC curve between HSR neurons and LSR neurons is quite similar for a 75 dB SPL stimulus. As such, it is expected that the JND estimation will not be too different between HSR and LSR neurons. This expectation is confirmed with Fig. 19 where the JND estimation for HSR neurons is similar to the JND estimation for LSR neurons.

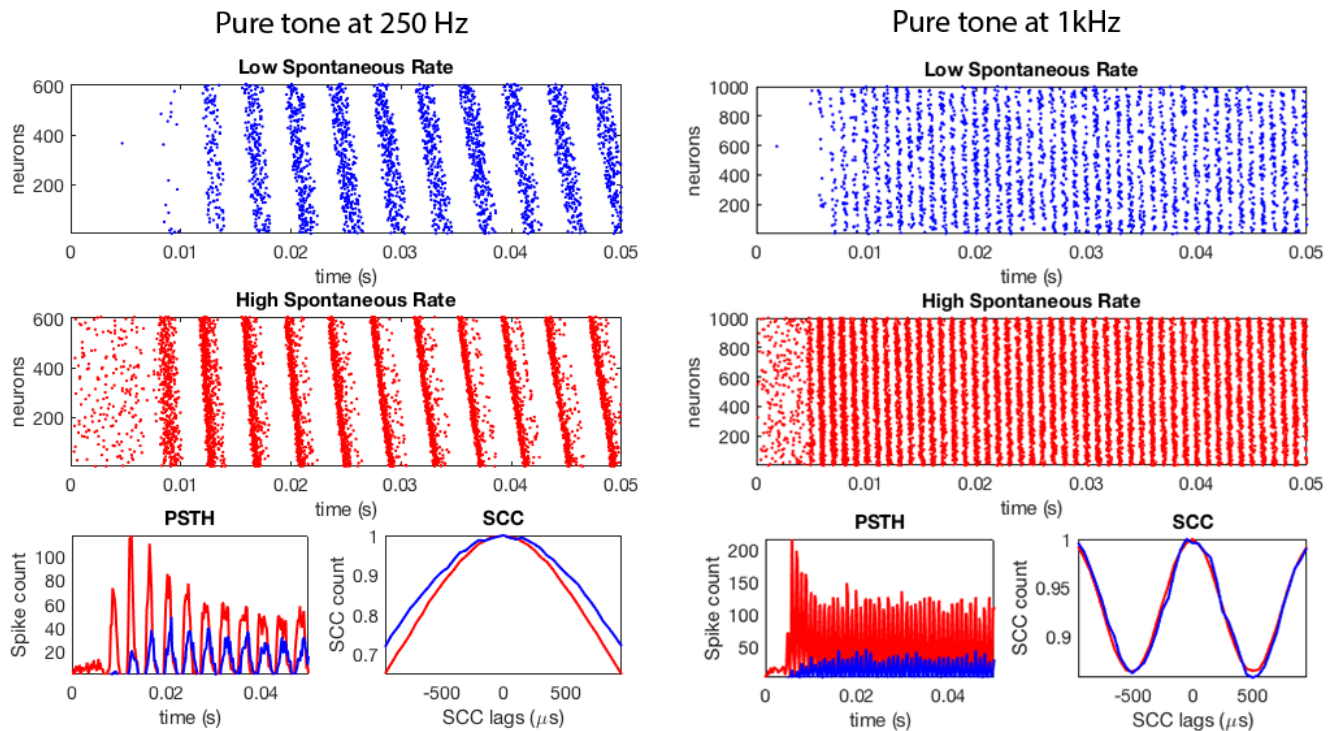


Figure 17: Pure tone signal at 75 dB SPL. Red is for high spontaneous rate and blue is for low spontaneous rate.

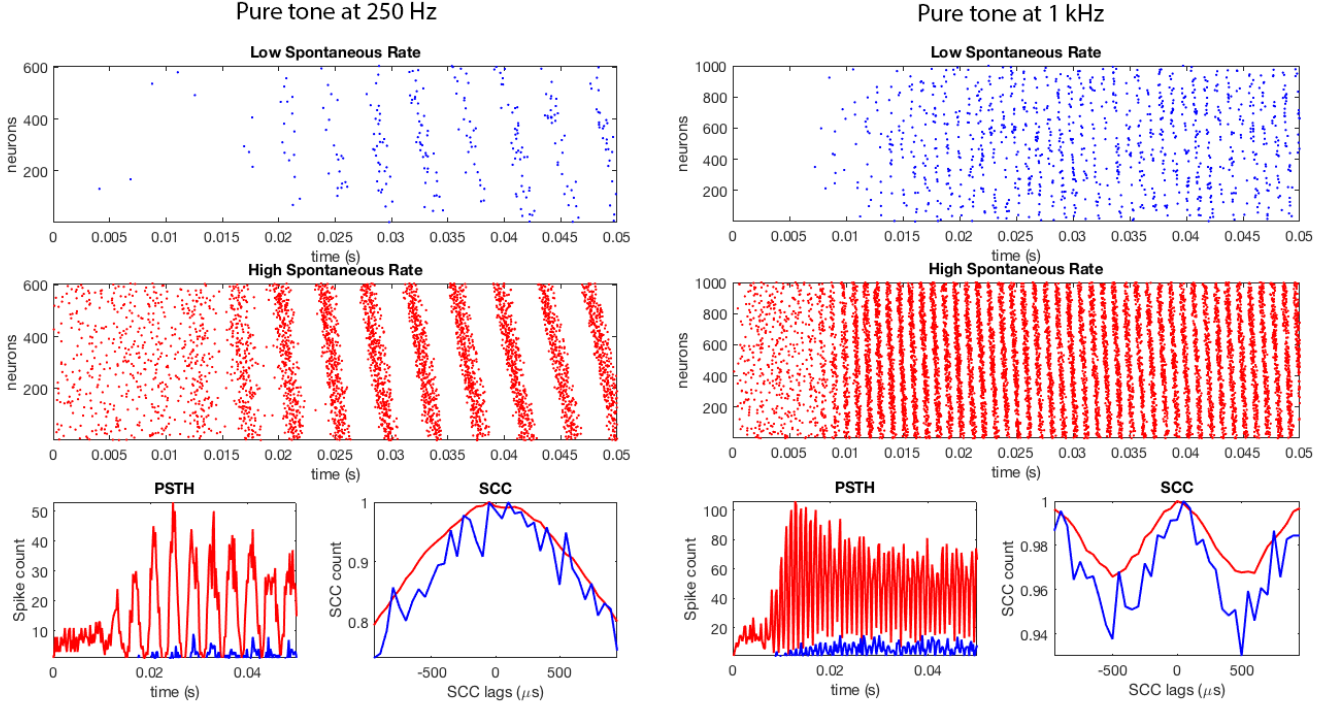


Figure 18: Pure tone signal at 40 dB SPL. Red is for high spontaneous rate and blue is for low spontaneous rate.

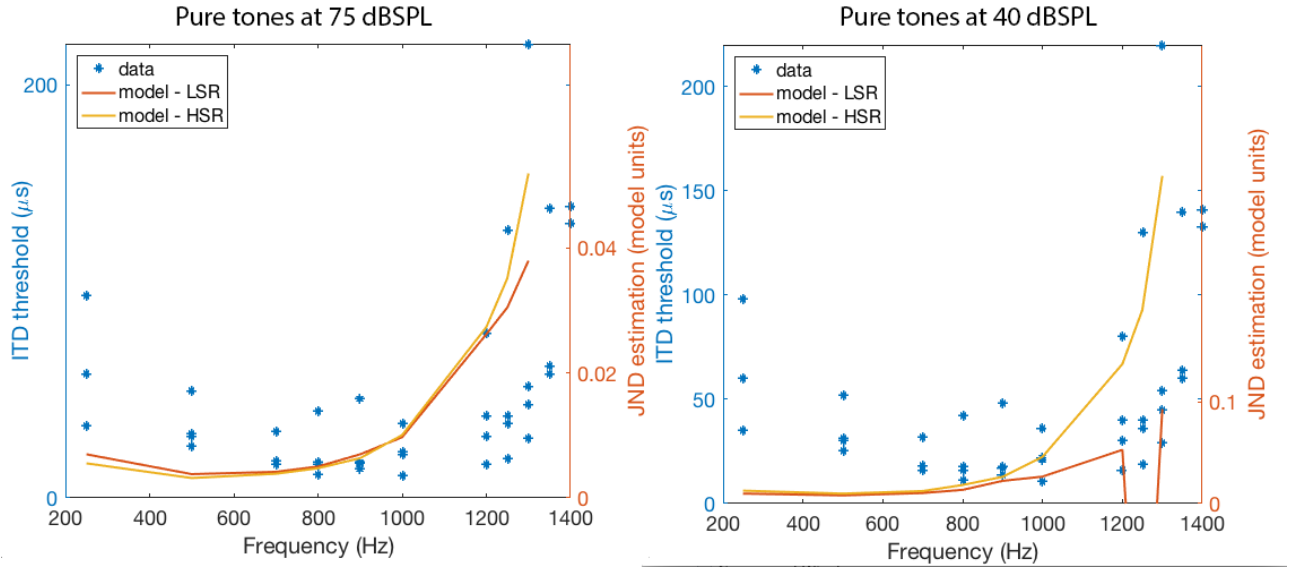


Figure 19: At 40 dB SPL the LSR spikes fail at describing the trend of the data, and both HSR and LSR fail to show change to lower frequencies (250 - 600 Hz). Also notice that the JND in model units is almost an order of magnitude greater for 40 dB SPL than 75 dB SPL.

Modulated stimuli

For the sake of brevity only the trapezoid stimulus (Laback et al., 2011) will be presented here, nonetheless similar results were observed for all the other modulated stimuli used.

In the previous section discussing the pure tone stimulus, the significance of the SCC curve shape was emphasised. Recall that the model normalises the SCC curve therefore

the average firing rate does not affect the outcome. As such, there is no significant difference between the HSR and LSR neuron response for the pure tone stimulus.

However, when considering modulated signals adaptation effects change the outcome. The portions of silent envelope appear to ‘restart’ the adaptive processes within the auditory nerve. This affects HSR neurons to a greater extent than LSR neurons, see Fig. 20. The effect of this on the model is that the SCC curves are sharper for HSR neurons than LSR neurons (Fig. 20), which leads to smaller JND values.

Furthermore, this lack adaptation ‘restart’ in the LSR neurons leads to wrong model predictions, compare Fig. 21 to Fig. 22. Specifically the experimental study by Laback et al. (2011) concluded that an increase in the trapezoid slope leads to lower JND, which is predicted using HSR neurons (see Fig. 21) whereas when using LSR neurons (see Fig. 22) the opposite effect is shown.

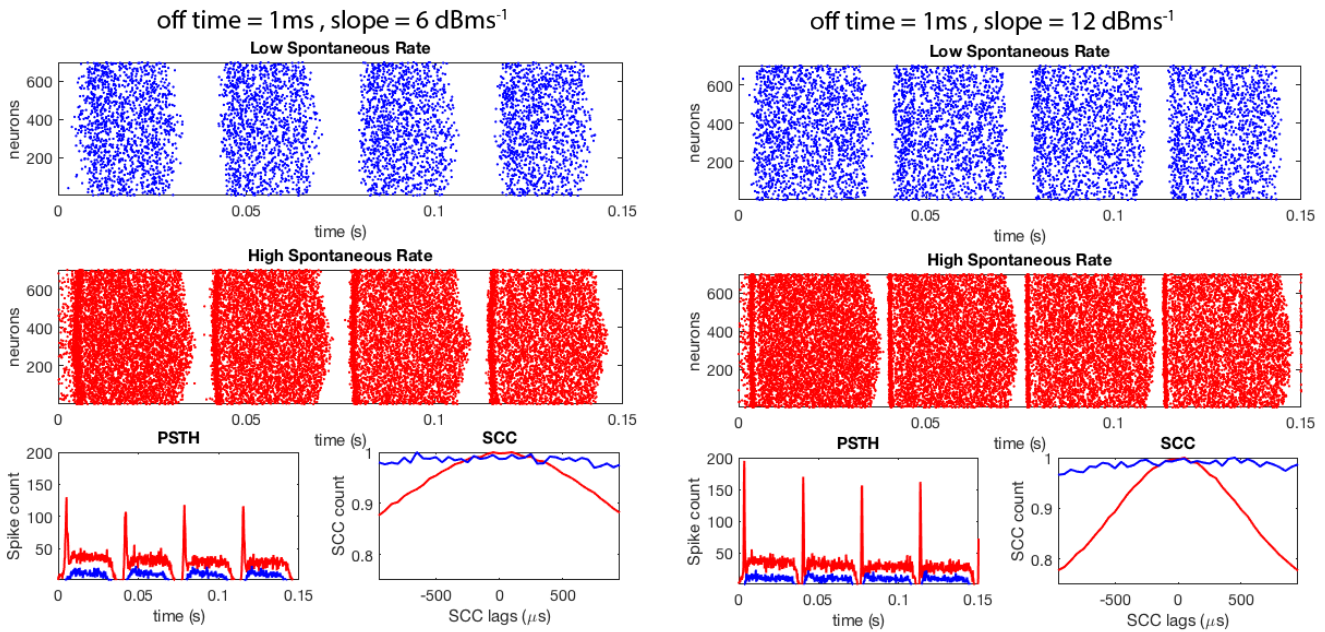


Figure 20: Trapezoid signal at 78.2 dB SPL. Red is for high spontaneous rate and blue is for low spontaneous rate.

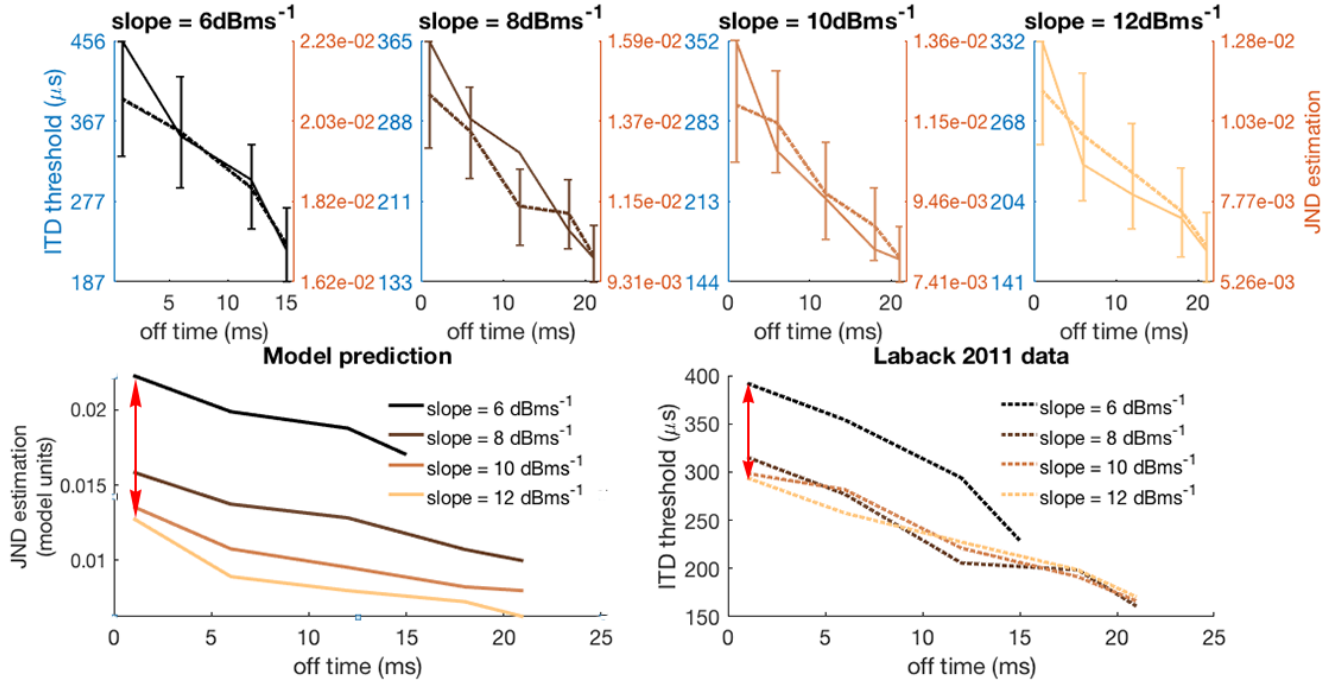


Figure 21: High Spontaneous Rate (HSR) neurons and their respective JND estimations. The red arrow indicates the two conditions shown in Fig. 20. Notice how the model agrees with the experimental study.

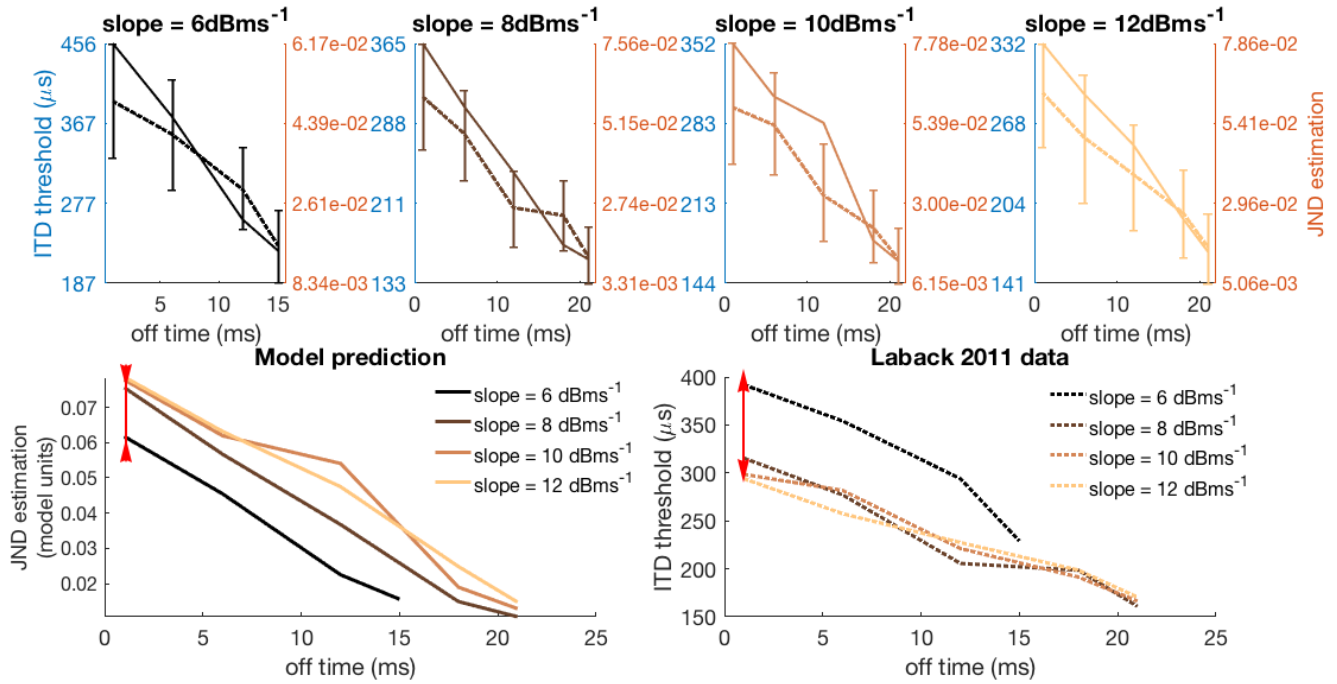


Figure 22: Low Spontaneous Rate (LSR) neurons and their respective JND estimations. The red arrow indicates the two conditions shown in Fig. 20. Notice how the model shows the opposite trend as the experimental study.

Mixed spontaneous rate neurons

Up to now only HSR neurons, or LSR neurons were used to simulate the auditory nerve response to sound stimuli. In human physiology however there is a mixture of HSR and LSR auditory neurons. The following figures show the model outcomes by using a mixture of 60% HSR neurons, 20% medium spontaneous rate (MSR) neurons and 20% LSR neurons¹.

As before, the major differences by the introduction of MSR and LSR neurons appear in the modulated waveforms, see the SCC peaks of Fig. 24. However, the differences between the two neurographs do not appear to be significant enough to change the final outcome of the model for the trapezoid tone analysis, see Fig. 25.

The effect observed from the inclusion of MSR and LSR neurons is a reduction in the response of neurons to the onset of modulating periods. This in turn affects the model outcome by estimating higher values for the JND in *model units*, but without altering the trends observed as the stimulus parameters are varied.

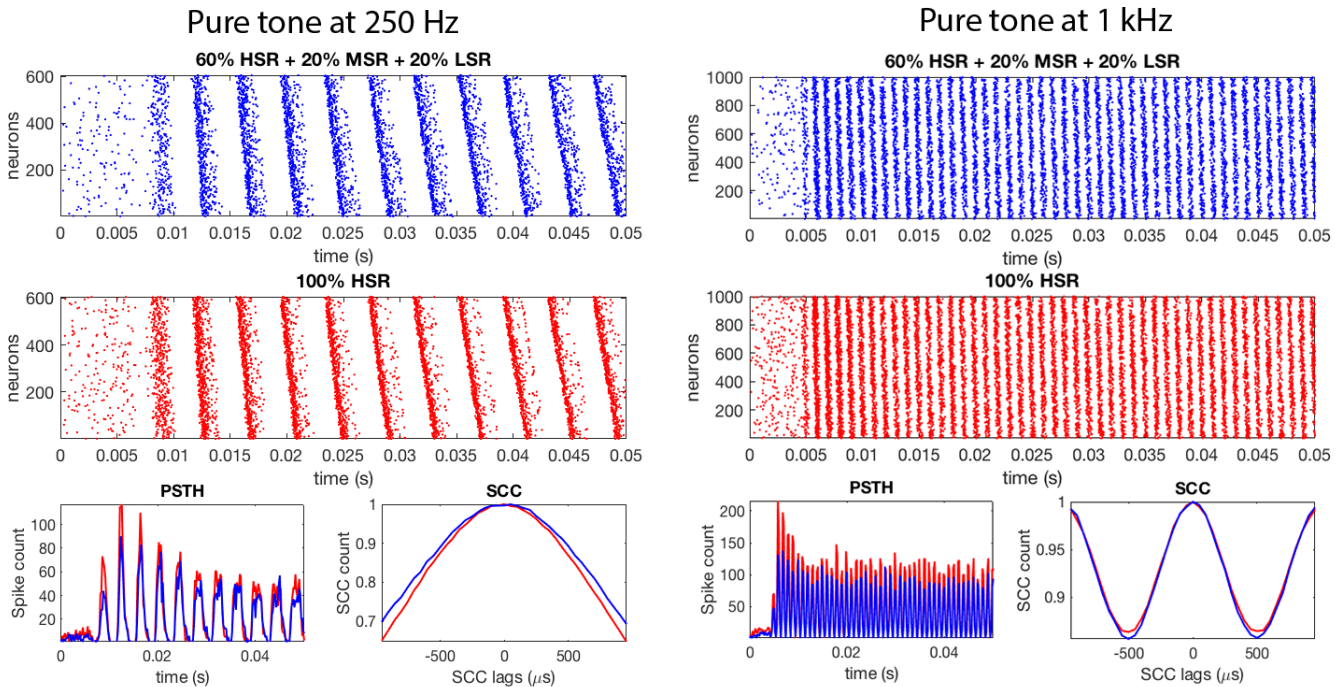


Figure 23: Pure tone at 75 dB SPL.

¹Liberman (1978) suggested the presence of 61% HSR, 23% MSR and 16% LSR neurons in the auditory nerve, notably of cats raised in a low sound exposure environment.

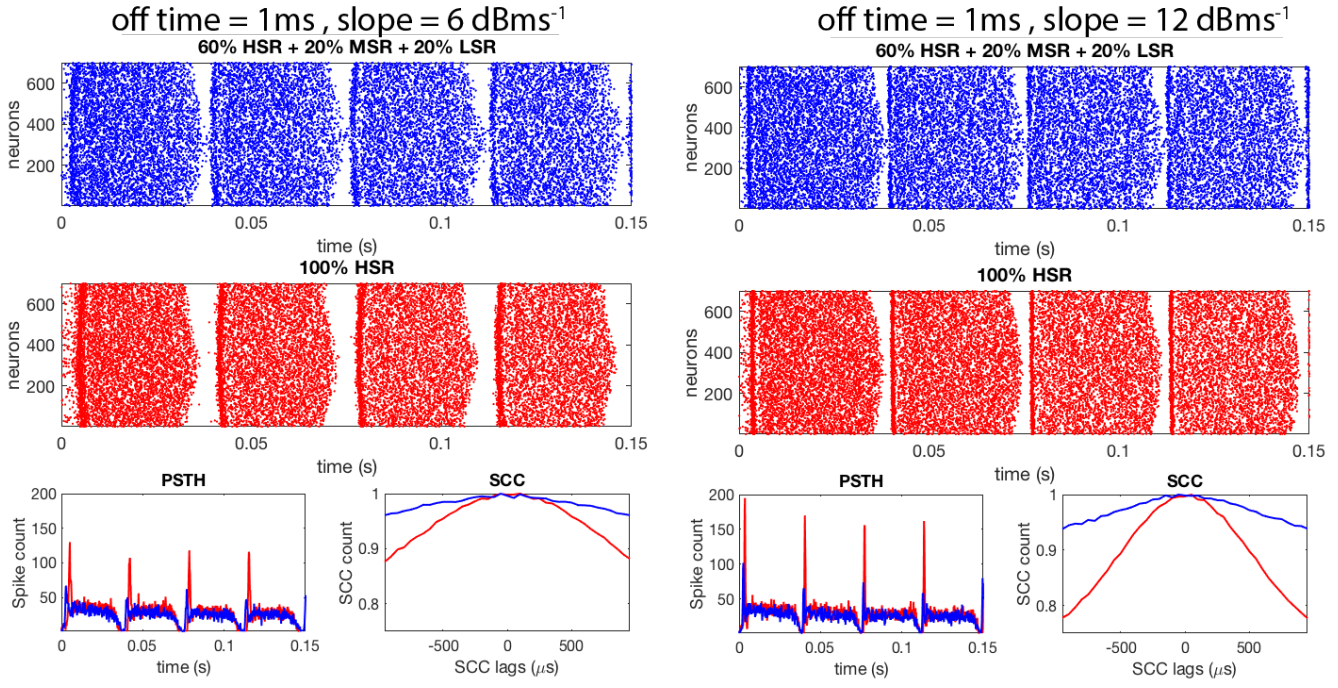


Figure 24: Trapezoidal modulated tone at 78 dB SPL.

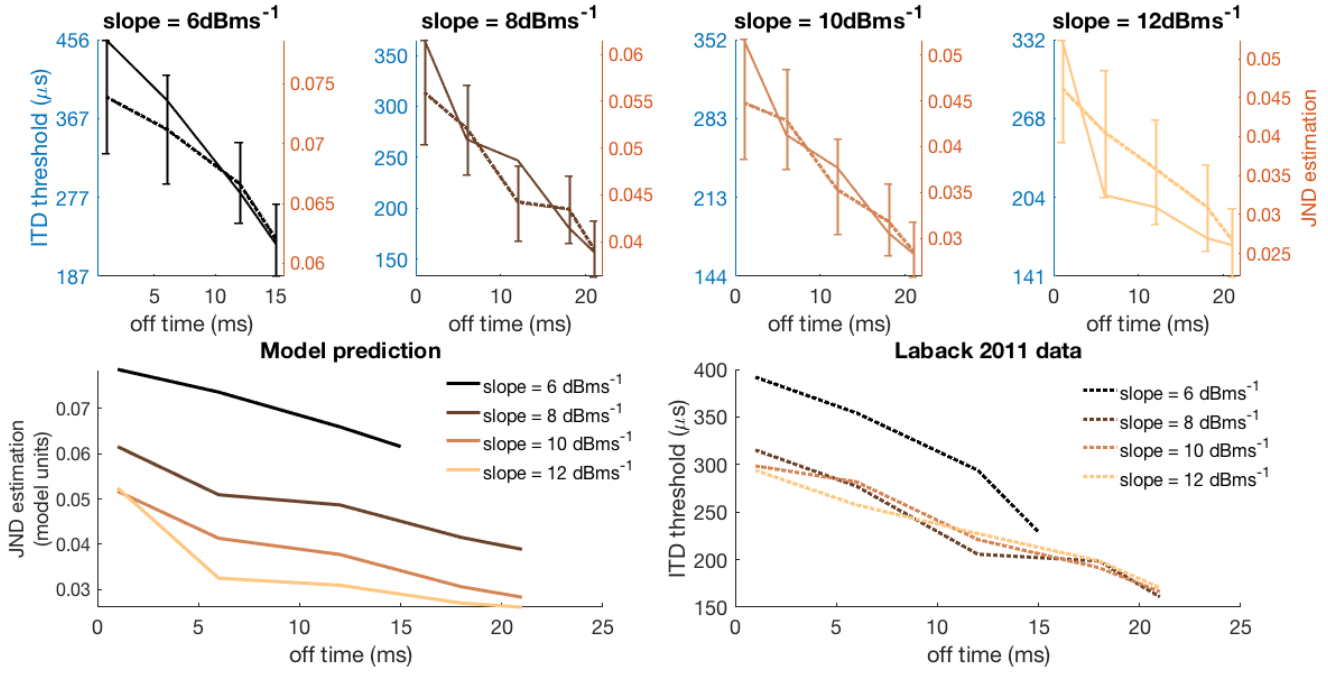


Figure 25: Trapezoid tones JND estimation using a mixture of 90% HSR neurons and 10% LSR neurons. Compare with Figure 11 of the manuscript.